Martínez Juan Fernando

HERRAMIENTAS ESTADÍSTICAS BÁSICAS PARA EL MEJORAMIENTO DE LA CALIDAD

HITOSHI KUME

Traducción
Eloisa Vasco

Revisión Técnica
Daniel Tarquino

GRUPO EDITORIAL norma

Bogotá, Barcelona, Buenos Aires, Caracas, Guatemala, Lima, México, Miami, Panamá, Quito, San José, San Juan, Santiago de Chile, Santo Domingo

Kume, Hitoshi
Herramientas estadísticas / Hitoshi Kume ; traducción Eloisa Vasco. —
Bogotá : Editorial Norma, 2002.
243 p. ; 10.6 x 18 cm.
ISBN 958-04-6719-6
Título original : Statistical methods for quality improvement.
1. Control de calidad – Métodos estadísticos I. Vasco, Eloisa, tr. II.
Tít.
658.568 cd 20 ed.
AHL5241

CEP-Banco de la República-Biblioteca Luis-Angel Arango

Edición original en inglés:
STATISTICAL METHODS FOR QUALITY IMPROVEMENT
de Hitoshi Kume, Ed.

Apartado Aéreo 53550, Bogotá, Colombia.
Edición revisada por España.
Gran Vía de les Corts Catalanas, 322-324, 08004 Barcelona.

Impreso por Ediciones Versalles Ltda.
Impreso en Colombia —Printed in Colombia

Dirección editorial, María del Mar Ravassa G.
Coordinación editorial, Cordillera Editores Ltda.
Editor, Oscar Torres D.
Diseñadora de cubierta, Marta Ayerbe

ISBN: 958-04-6719-6



Contenido

Prefacio ix

Capítulo I
Introducción 1
El papel de los métodos estadísticos en la administración de los procesos de producción 1

Capítulo II
Cómo obtener datos 8
2.1 Cómo recoger datos 8
2.2 Las hojas de registro 11

Capítulo III
El análisis de Pareto 19
3.1 ¿Qué son los diagramas de Pareto? 19
3.2 Cómo elaborar diagramas de Pareto 20
3.3 Diagramas de Pareto de fenómenos y diagramas de Pareto de causas 23
3.4 Notas sobre los diagramas de Pareto 24

Capítulo IV
Diagramas de causa-efecto 27
4.1 ¿Qué son los diagramas de causa-efecto? 27
4.2 Cómo elaborar diagramas de causa-efecto 28
4.3 Notas sobre los diagramas de causa-efecto 33
4.4 Diagramas de Pareto y diagramas de causa-efecto 34



Capítulo V
Los histogramas 39
5.1 Distribuciones e histogramas 39
5.2 Cómo elaborar histogramas 42
5.3 Cómo leer histogramas 50
5.4 Medidas para presentar las características de las distribuciones 55
5.5 La distribución normal y sus características 61

Capítulo VI
Los diagramas de dispersión 68
6.1 ¿Qué son los diagramas de dispersión? 68
6.2 Cómo elaborar un diagrama de dispersión 69
6.3 Cómo leer los diagramas de dispersión 72
6.4 El cálculo de los coeficientes de correlación 75
6.5 Notas sobre el análisis correlacional 78
6.6 ¿Qué es el análisis de regresión? 83
6.7 Estimación de las líneas de regresión 85
6.8 Notas sobre el análisis de regresión 87

Capítulo VII
Gráficas de control 91
7.1 ¿Qué son las gráficas de control? 91
7.2 Tipos de gráficas de control 93
7.3 Cómo elaborar una gráfica de control 96
7.4 Cómo leer las gráficas de control 105
7.5 Análisis del proceso usando las gráficas de control 108
7.6 Estudio de caso de análisis del proceso 115
7.7 Control del proceso con gráficas de control 134

Capítulo VIII
Aditividad de las varianzas 143
8.1 Las medias y las varianzas de sumas 143
8.2 Precisión del ensamblaje de partes 146
8.3 Fórmulas teóricas 148
8.4 El valor esperado y la varianza de la media muestral . 149
8.5 Error de muestreo y error de medición 149
8.6 La varianza de los valores de una función 151
8.7 Cuando las variables aleatorias no son independientes 152



8.8 Combinación selectiva 154
8.9 Control estadístico de calidad 155

Capítulo IX
Introducción a la inferencia estadística 158
9.1 Estadística 158
9.2 Distribución de las estadísticas 159
9.3 Prueba de la hipótesis 169
9.4 Estimación de parámetros 177
9.5 Pruebas y estimaciones de las medias de poblaciones cuando xxx no se conoce 178
9.6 Pruebas y estimaciones de las diferencias entre las medias de dos poblaciones 182
9.7 Pruebas y estimaciones en observaciones pareadas 185
9.8 Pruebas de significación de los coeficientes de correlación 189

Capítulo IX
La ruta de la calidad 193
10.1 El problema 195
10.2 Observación 197
10.3 Análisis 201
10.4 Acción 206
10.5 Verificación 207
10.6 Estandarización 208
10.7 Conclusión 210

Epílogo 212

Apéndice 214
Tabla A.1 Tabla de la distribución normal 214
Tabla A.2 Coeficientes para la gráfica $\overline{x}$—R 214
Tabla A.3 Puntos de porcentaje para la distribución t 214

Respuestas a los ejercicios 218

Índice 193

Prefacio

Los métodos estadísticos son herramientas eficaces para mejorar el proceso de producción y reducir sus defectos. Sin embargo, se debe tener en cuenta que las herramientas estadísticas son precisamente herramientas: no servirían si se usan inadecuadamente.

Con frecuencia se intenta reducir los defectos de producción remontándose directamente a la causa del defecto. Ése es un enfoque directo y, a primera vista, parece que es eficiente. Pero, en la mayoría de los casos, las causas encontradas por medio de ese enfoque no son las verdaderas. Si se aplican soluciones a los defectos basándose en el conocimiento de esas causas falsas, el intento puede no tener resultados y el esfuerzo se perderá. El primer paso para encontrar la verdadera causa es una observación cuidadosa del fenómeno del defecto. Luego de esa observación cuidadosa, la verdadera causa será evidente.

Las herramientas estadísticas dan objetividad y precisión a las observaciones. Las premisas de la manera de pensar estadística son:

1) Dele mayor importancia a los hechos que a los conceptos abstractos.
2) No exprese los hechos en términos de sentimientos o de ideas. Utilice cifras derivadas de los resultados específicos de la observación.



3) Los resultados de las observaciones, acompañados como están por el error y la variación, son parte de un todo oculto. Encontrar ese todo oculto es la finalidad última de la observación.
4) Acepte como información confiable, la distribución normal que aparece cuando hay un gran número de observaciones.

En primer lugar, se debe reconocer la imperfección del reconocimiento humano. Después debe entenderse que el conocimiento actual no es más que la base para nuevas hipótesis. Sabiendo esto, los métodos de pensamiento mencionados antes pueden ser útiles para profundizar nuestro entendimiento del proceso de producción y de las formas de mejorarlo.

Este libro difiere de los textos de estadística comunes. Su propósito es mostrar cómo aplicar los métodos estadísticos a los problemas del mundo real. Quienes tengan poco conocimiento del análisis estadístico también se beneficiarán. Debe comenzar con los pasos iniciales de la explicación y después seguir con los otros capítulos. Las personas que ya conozcan la estadística pueden omitir los pasos introductorios y empezar con las explicaciones del análisis estadístico en su uso práctico.

Los autores se sentirán muy satisfechos si el libro ayuda a los gerentes de producción y a los ingenieros a mejorar el proceso de producción.

Los capítulos son revisiones de artículos que aparecieron por primera vez en la revista *Kenshu*, publicada por la Asociación de Becas Técnicas Internacionales (Association for Overseas Technical Scholarships, AOTS), una asociación sin ánimo de lucro. El objetivo de AOTS es promover la cooperación técnica. Los lectores de *Kenshu* son gerentes de producción e ingenieros de países en vía de desarrollo. En seguida se enuncian los capítulos del libro junto con sus respectivos autores, pero la responsabilidad por los errores y las omisiones es del editor.

Capítulo	1	Introducción	Hitoshi Kume
Capítulo	2	Cómo obtener datos	Yoshinori Iizuka



I

Introducción

EL PAPEL DE LOS MÉTODOS ESTADÍSTICOS EN LA ADMINISTRACIÓN DE LOS PROCESOS DE PRODUCCIÓN

(1) ¿Qué causa los productos defectuosos?

Uno tras otro, los productos llegan en la banda transportadora. Al final de la banda transportadora hay una máquina empacadora que continuamente empaca los productos que llegan y los envía a la bodega. Una mirada más atenta permite ver a un hombre de pies entre la banda transportadora y la máquina empacadora. Observa cuidadosamente los productos que llegan, y ocasionalmente recoge algunos y los arroja casualmente al cesto que está detrás de él. Ésos son los productos defectuosos.

Este tipo de cosas se ve comúnmente en muchas fábricas. En un principio, esos productos descartados parecen desperdiciados, pero pronto se aceptan como una parte del proceso. Pero acostumbrarse a los productos defectuosos no soluciona el problema, sino que más bien constituye un retroceso en la solución.

¿Cómo se producen los productos defectuosos? ¿Qué debe hacerse para reducir su número?

Para poder disminuir el número de productos defectuosos es necesario creer que sí se puede reducir. Sobra decir que la sola creencia no disminuirá el número de productos defectuosos. Lo que queremos decir es que existen causas particulares para que un producto dado resulte defectuoso, y que los productos defectuosos pueden desaparecer si se descubren y se eliminan esas causas.

Mucha gente piensa que debido a los estrictos requisitos de calidad que deben cumplir sus productos y a los muchos factores que pueden causar un defecto, los productos defectuosos son inevitables.

Sin embargo, independientemente de los tipos de productos o de las clases de métodos de producción utilizados, la causa de los defectos es universal.

La variación: Ésta es la causa. ¿Qué sucedería si fabricáramos productos usando materiales de exactamente la misma calidad, máquinas y métodos de trabajo idénticos y si inspeccionáramos estos productos exactamente de la misma manera? No importa cuántos productos se fabriquen, todos ellos deben ser idénticos mientras las cuatro condiciones anteriores sean idénticas. Es decir, todos los productos cumplirán o no cumplirán los requisitos. Todos ellos serán defectuosos si los materiales, la maquinaria, el método de trabajo o la inspección son inadecuados. En este caso, se producirán productos defectuosos exactamente idénticos. Mientras no haya fallas en las cuatro condiciones mencionadas, los productos que resulten deben ser todos "idénticamente" productos no defectuosos.

Con respecto a los productos que fabricamos, es casi imposible que todos ellos salgan defectuosos. Algunos son defectuosos mientras que otros no lo son. En otras palabras, los productos defectuosos y los buenos salen mezclados.

¿Por qué se producen en un mismo lote productos defectuosos y no defectuosos? La causa, como se ha dicho antes, es la variación. La variación en *materiales*, en las *condiciones de la má-*



quina, en los *métodos de trabajo* y en las *inspecciones* es la causa de los productos defectuosos. Si no existiera ninguna de estas variaciones, todos los productos serían idénticos y no habría variaciones en calidad, tales como la ocurrencia de productos defectuosos y no defectuosos.

Pensemos en el trabajo de doblar planchas de acero. Todas las planchas de acero parecen tener el mismo espesor, pero cuando se miden con precisión, tendrán espesores diferentes. Además, aun la misma plancha tendrá más espesor en unas partes que en otras. Si vamos más lejos e inspeccionamos la estructura de cristal de las planchas, habrá pequeñas variaciones en la forma de los cristales compuestos de hierro, carbono y otros elementos, de una a otra parte de la plancha. Estas diferencias afectan naturalmente a las características de calidad. Aun usando el mismo método, las planchas no se doblarán de manera uniforme. Algunas de ellas podrán incluso rajarse.

Examinemos ahora la maquinaria. La cortadora pierde filo a medida que procesa cierto número de productos. Las condiciones del aceite lubricante también cambian con los cambios en temperatura. Las dimensiones de los productos cambian según la forma como se ajuste el mecanismo de corte. Aunque pareciera que una operación se realiza exactamente en las mismas condiciones que otra, muchos cambios o variaciones ocurren sin que se noten, y afectan a la calidad del producto.

Pensemos ahora en el tratamiento con calor. La temperatura de los hornos cambia continuamente con los cambios de voltaje en el caso de un horno eléctrico, y con los cambios en la presión del gas en el caso de un horno a gas. Dentro del horno, las áreas cerca de la boca, el techo, el piso o la pared del horno tienen diferentes condiciones. Cuando introducimos en el horno materiales para tratamiento por calor, la cantidad de calor que reciben los materiales varía según su posición relativa, afectando a características de calidad, tales como la dureza del producto.

Las características físicas de los trabajadores y su habilidad también afectan a la variación de la calidad de los productos. Hay hombres altos y bajos, hábiles y menos hábiles, hombres



con recia musculatura y hombres débiles, personas diestras y zurdas. Todos los trabajadores podrán creer que están trabajando de la misma manera, pero hay diferencias personales. Aun el mismo individuo trabaja de manera diferente según como se sienta ese día particular y según sus condiciones de fatiga. A veces puede cometer un error por descuido.

En la inspección, pueden ocurrir variaciones de calidad aparentes. Si se usa un calibrador en la inspección, pueden causarse variaciones en los datos por un mal calibrador o por la forma de usarlo. En el caso de inspecciones sensoriales, como la inspección visual, la calidad parece variar si hay variación en el criterio del inspector. La variación en la inspección no se relaciona directamente con la calidad del producto, pero afecta al proceso de decidir si un producto es defectuoso o no lo es.

Si analizamos el problema de esta manera, podemos ver que en el proceso de fabricación de un producto intervienen innumerables factores que afectan a las características de calidad de ese producto. Cuando consideramos el proceso de manufactura desde el punto de vista de la variación de la calidad, podemos pensar en el proceso como *un agregado de las causas de la variación*. Estas causas son la explicación de los cambios en las características de calidad de los productos, que tienen como resultado productos defectuosos o no defectuosos. Un producto se considera no defectuoso si las características de calidad satisfacen cierto requisito, y defectuoso si no lo hacen. Por lo tanto, aun los productos no defectuosos tienen variaciones dentro del mismo requisito. Esto quiere decir que no existen los productos "exactamente iguales" de los que hablamos antes.

Los productos defectuosos son causados por las variaciones. Si estas variaciones se reducen, seguramente disminuirán los productos defectuosos. Éste es un principio sencillo y sólido, aplicable cualesquiera que sean los tipos de productos o las clases de métodos de producción utilizados.



(2) El diagnóstico de los procesos

Aunque las causas de la variación en la calidad son innumerables, no toda causa afecta a la calidad en el mismo grado. Algunas la afectan enormemente, mientras que otras, aunque teóricamente consideradas como muy importantes, tienen poco efecto sobre la variación en la calidad cuando se controlan adecuadamente.

Las innumerables causas concebibles pueden categorizarse en dos grupos, el primero de los cuales consiste en un pequeño número de causas que, sin embargo, tienen un gran efecto (los *pocos vitales*), y un segundo grupo que incluye muchas causas que tienen sólo efectos menores (los *muchos triviales*). Generalmente, no hay muchos factores que realmente causen defectos. Este hecho se llama *principio de Pareto* y se usa en muchos casos.

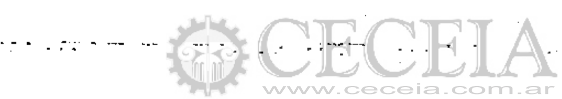
Con la aplicación del ya mencionado principio de variación y de éste de Pareto se facilita considerablemente el problema de reducir el número de productos defectuosos. Lo que necesitamos hacer es encontrar las pocas causas vitales de los productos defectuosos y eliminar estas causas después de que se hayan identificado claramente. "En nuestro proceso hay tantas causas de productos defectuosos que es realmente imposible controlarlas". Comentarios como éste se escuchan con frecuencia en sitios donde los procesos están llenos de productos defectuosos. Todo proceso tiene muchas causas de variación en la calidad, y ningún proceso tiene un número especialmente alto de esas causas. Hay una gran diferencia entre tener muchos "sospechosos" que pueden estar causando los defectos y realmente tener muchos "culpables" que de hecho estén causando los defectos.

El proceso de encontrar las causas de los productos defectuosos entre muchos factores se llama *diagnóstico del proceso*. Para reducir el número de productos defectuosos, la primera acción necesaria es hacer un diagnóstico correcto para ver cuáles son las verdaderas causas de los defectos. Si esto no se hace correctamente, no se puede reducir el número de productos defectuosos. Es como dar a un paciente de apendicitis un remedio para la diges-

ción, que, desde luego, no lo cura. El efecto puede ser que el paciente se sienta mejor por un tiempo, pero después la enfermedad recurrirá en una forma peor que antes.

¿Cómo hacer un diagnóstico correcto? Hay muchos métodos. Algunos utilizan la intuición, otros dependen de la experiencia. Aun otros recurren al análisis estadístico de los datos, y hasta se puede usar la investigación experimental. El método intuitivo se usa con frecuencia porque es muy rápido. De hecho, hay algo más que la intuición ordinaria en la intuición de un verdadero experto, y debe respetarse. Un movimiento intuitivo hecho por un experto jugador de ajedrez es superior al movimiento hecho por cien aficionados. El consejo y la intuición de los especialistas y los expertos debe respetarse profundamente. Sin embargo, la dificultad en el problema de reducir el número de productos defectuosos es que no siempre es claro quién es el verdadero experto. En el caso del ajedrez, el consejo de los expertos merece casi total confianza porque tanto el fuerte como el débil se manifiestan en juegos reales, y los campeones son aquéllos que han ganado y han sobrevivido a duros torneos. En el caso del diagnóstico de proceso, sin embargo, es frecuente que quien parece ser un buen "médico" no lo sea necesariamente y puede resultar que sea alguien que "haya dejado morir muchos pacientes". Además, en una época de progreso rápido, es difícil seguir siendo un experto en todos los problemas cuya naturaleza está cambiando constantemente. Debido a que el problema de los productos defectuosos se encuentra con frecuencia en áreas en las cuales no hay experiencia previa, lo que se necesita no es muchos años de experiencia sino la decisión de reducir el número de productos defectuosos y una actitud de observación de la situación real en forma objetiva. La forma estadística de considerar las cosas y el uso de los métodos estadísticos es un medio muy eficaz para hacer esta observación.

Los métodos estadísticos proporcionan un medio eficaz para desarrollar una nueva tecnología y controlar la calidad de los procesos de manufactura. Muchas empresas manufactureras importantes han estado tratando de usar activamente los métodos esta-



dísticos, y algunas han dedicado más de 100 horas al año a educación interna sobre este tema. El conocimiento de los métodos estadísticos se ha convertido en parte normal de la capacitación de un ingeniero, pero el conocimiento de los métodos estadísticos no proporciona inmediatamente la habilidad para usarlos. La habilidad para analizar las cosas desde el punto de vista estadístico es más importante que los métodos individuales. Además, necesitamos ser francos para reconocer los problemas y la variación, y recoger información sobre ellos. Finalmente, queremos subrayar que lo importante no es solamente el conocimiento de los métodos estadísticos como tales sino más bien la actitud mental hacia su utilización.

II

Cómo obtener datos

2.1 CÓMO RECOGER DATOS

(1) Establezca objetivos claros

La información es una guía para nuestras acciones. A partir de la información conocemos los hechos pertinentes y adoptamos acciones apropiadas basadas en esos hechos. Antes de recoger la información, es importante determinar qué se va a hacer con ella.

En una fábrica constructora de máquinas, se le hace un muestreo de inspección de calidad a cierto tipo de accesorio comprado a un proveedor. Se dio el caso de que un lote que debería haberse rechazado se aceptó como excepción especial, para poder cumplir el programa de producción. Sin embargo, no se hizo nada especial respecto al lote que había sido aceptado. Esto quiere decir que tanto los lotes que se ajustaban a las especificaciones como aquéllos que no se ajustaban pasaron al proceso siguiente. En realidad, estos datos se estaban obteniendo para determinar la aceptabilidad de los lotes, pero no se usaron para nada.

En el control de calidad, los objetivos de la recolección de información son:



1) El control y el monitoreo del proceso de producción.
2) El análisis de lo que no se ajusta a las normas.
3) La inspección.

Cualquier recolección de información ha de tener un propósito específico y ser seguida por acciones.

(2) ¿Cuál es su propósito?

Una vez que se define el objetivo de la recolección de información, también se determinan los tipos de comparación que se necesitan, y esto a su vez identifica el tipo de datos que se deben recoger. Por ejemplo, suponga que hay una pregunta respecto a la variación en una característica de calidad de un producto. Si solamente se recoge un dato cada día, será imposible determinar la variación ese día. O, si usted quiere saber por qué resultan productos defectuosos hechos por dos trabajadores diferentes, es necesario tomar las muestras separadamente para poder comparar el desempeño de cada uno de ellos. Si la comparación entre ellos muestra una clara diferencia, una medida remedial que elimine la diferencia entre los trabajadores reducirá también la variación en el proceso.

Esta división de un grupo en varios subgrupos, con base en ciertos factores, se llama *estratificación*. La estratificación es muy importante, y es necesario que su aplicación se convierta en un hábito de pensamiento en todo tipo de situaciones.

Suponga que usted quiere saber la relación entre la calidad de un ingrediente y la dureza del producto. En un caso como éste, cuando usted quiere saber si hay una relación entre los valores de dos características, los datos tienen que estar disponibles por pares. Si los datos se recogen por pares, se pueden analizar usando un *diagrama de dispersión,* que se explica en el capítulo VI.

(3) ¿Son confiables las mediciones?

Incluso si las muestras se han seleccionado adecuadamente, se hará un juicio erróneo si las mediciones no son confiables. Por ejemplo, las inspecciones hechas por cierto inspector mostraron que una fracción de productos defectuosos era muy diferente de las demás, y un examen cuidadoso mostró más tarde que un instrumento de medición se había descompuesto.

En el caso de mediciones sensoriales, tales como la inspección visual, las diferencias debidas a los inspectores individuales son comunes. Este hecho debe tenerse en cuenta cuando se recogen y se analizan datos.

(4) Establezca formas apropiadas de recoger los datos

Una vez que se han recogido los datos, diferentes clases de métodos estadísticos pueden ser utilizados para analizarlos, de modo que se conviertan en fuente de información. Cuando se recogen datos, es importante organizarlos adecuadamente para facilitar su procesamiento posterior. En primer lugar, el origen de los datos debe registrarse claramente. Los datos cuyo origen no se conoce con claridad se convierten en información inútil. Con frecuencia, se obtiene poca información útil a pesar de haber gastado una semana reuniendo datos sobre alguna característica de calidad, debido a que las personas olvidaron en qué días de la semana se recogieron los datos, qué máquinas hicieron el proceso, quiénes fueron los trabajadores, qué lotes de materiales se usaron, y así sucesivamente.

En segundo lugar, los datos deben registrarse de tal manera que puedan utilizarse fácilmente. Por el hecho de que con frecuencia los datos se utilizan posteriormente para cálculos estadísticos, tales como promedios y rangos, es mejor registrarlos de tal manera que estos cálculos se faciliten. Por ejemplo, los datos sobre 100 piezas, obtenidos haciendo mediciones cuatro veces al día (a las 9:00, 11:00, 2:00 y 4:00) durante 25 días, normalmente se registran en una hoja de datos, como la que se muestra en la tabla 2.1, en la cual la hora se organiza horizontalmente y los



Fecha	Hora			
	9 a.m.	11 a.m.	2 p.m.	4 p.m.
Feb. 1	12.3	11.5	13.2	14.2
Feb. 2	13.2	12.5	14.0	14.0
Feb. 3	....	....	....	....

Tabla 2.1 Un ejemplo de una hoja de datos

días verticalmente. De esta manera, los cálculos diarios pueden hacerse marcando los números en cada renglón, y los cálculos para las horas pueden hacerse dentro de cada columna. Cuando se necesita registrar datos de manera continua se recomienda preparar formatos para su registro.

2.2 LAS HOJAS DE REGISTRO

Como se dijo en la sección anterior, si se llega a la conclusión de que es necesario reunir información, es esencial que el objetivo sea claro y que se obtengan datos que reflejen los hechos con claridad. Además de estos requisitos, en situaciones reales es importante que los datos se recojan en forma clara y fácil de usar. Una *hoja de registro* es un formato preimpreso en el cual aparecen los ítems que se van a registrar, de tal manera que los datos puedan recogerse fácil y concisamente. Sus objetivos principales son dos:

1) Facilitar la recolección de los datos.
2) Organizar automáticamente los datos de manera que puedan usarse con facilidad más adelante.

La recolección y el registro de los datos parece fácil pero en realidad es difícil. Generalmente, mientras más personas procesen los datos, mayor es la probabilidad de que se presenten erro-

res de transcripción. Por lo tanto, la hoja de registro, en la cual los datos puedan registrarse por medio de cruces o de símbolos sencillos y en la cual los datos se organizan automáticamente sin necesidad de más copias a mano, se convierte en una herramienta poderosa para el registro de los datos. A continuación se presentan algunos ejemplos de hojas de registro.

Ejemplo 2.1
Hoja de registro para la distribución del proceso de producción

	Desv.	Registros																				Frecuencia
						5					10					15					20	
	−10																					
	−9																					
Especificación	−8																					
	−7																					
	−6																					
	−5	X																				1
	−4	X	X																			2
	−3	X	X	X	X																	4
	−2	X	X	X	X	X	X															6
	−1	X	X	X	X	X	X	X	X	X												9
8.300	0	X	X	X	X	X	X	X	X	X	X	X										11
	1	X	X	X	X	X	X	X	X													8
	2	X	X	X	X	X	X	X														7
	3	X	X	X																		3
	4	X	X																			2
	5	X																				1
	6	X																				1
	7																					
Especificación	8																					
	9																					
	10																					
																		Total				55

Figura 2.1 Hoja de registro para ver la distribución de un proceso de producción



Suponga que queremos saber las variaciones en las dimensiones de cierta clase de partes cuya especificación de maquinado es 8.300 ± 0.008. Para estudiar la distribución de los valores característicos del proceso, se usan normalmente los histogramas. Valores tales como la media y la varianza se calculan con base en el histograma y también se examina de diversas maneras la forma de la distribución.

Cuando se elabora un histograma, es doble trabajo recoger una gran cantidad de datos y después hacer una gráfica que muestre la distribución de frecuencias. Una manera más sencilla es clasificar los datos exactamente en el momento de recogerlos. La figura 2.1 es un ejemplo de un formato que puede prepararse con anterioridad. Se coloca una cruz en la casilla correspondiente cada vez que se realiza una medición, de manera que el histograma esté listo cuando se acaben de hacer las mediciones. Cuando se requiera estratificar y se use una sola hoja de registro, es mejor usar diferentes colores o símbolos, de manera que la diferencia pueda reconocerse más tarde.

Ejemplo 2.2
Hoja de registro de ítems defectuosos

La figura 2.2 muestra una hoja de registro utilizada en la inspección final de cierto producto de plástico moldeado. El inspector hace una marca en el registro cada vez que encuentra un defecto. Al finalizar el día de trabajo puede calcular de inmediato el número total y el tipo de defectos que ocurrieron.

El simple conocimiento del número total de los defectos no lleva a acciones correctivas, pero si se usa una hoja de registro como la de esta figura, podrán obtenerse indicios muy importantes para el mejoramiento del proceso porque la información muestra claramente cuáles son los defectos más frecuentes y cuáles no.

Si se usa esta hoja de registro, lo mismo que ocurre con el ejemplo 2.1, será imposible estratificar los datos más tarde, por ejemplo, según se hayan recogido en la mañana o en el tarde. Por

Hoja de registro

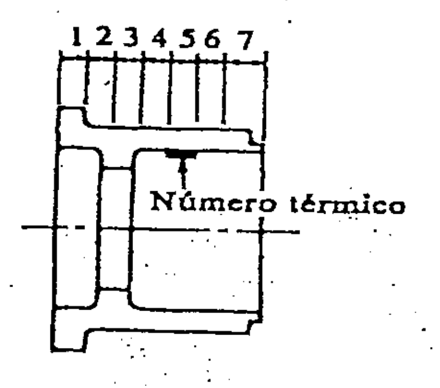
Producto:	Fecha:
Etapa de manufactura: Insp. final	Sección:
Tipo de defecto: rayones, incompleto, rajado, deformado	Nombre del inspector:
Número total inspeccionado: 1.525	Número del lote:
Observaciones: Se inspeccionaron todos los ítems	Número de orden:

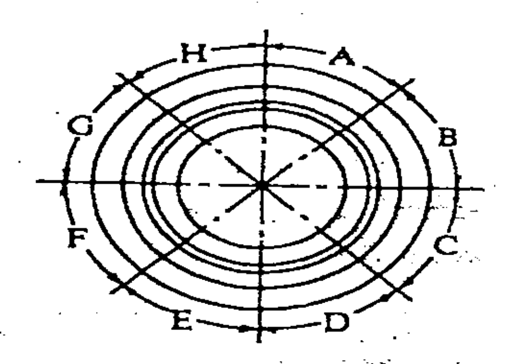
Tipo	Registro	Subtotal
Rayas superficiales	////\ ////\ ////\ //	17
Rajaduras	////\ ////\ /	11
Incompleto	////\ ////\ ////\ ////\ ////\ /	26
Deforme	///	3
Otros	////\	5
	Total:	62
Total rechazados	////\ ////\ ////\ ////\ ////\ ////\ ////\ ////\ //	42

Figura 2.2 Hoja de registro de ítems defectuosos

lo tanto, si se considera que la estratificación es necesaria, se debe incorporar en el diseño desde el comienzo.

Es necesario decidir con claridad de antemano cómo registrar los defectos cuando se encuentren dos en un mismo producto, y dar instrucciones precisas y completas a las personas que van a hacer el registro. En el caso de la figura 2.2, 42 de las 1.525 piezas estaban defectuosas. Sin embargo, el número total de de-





2. Matriz de localización del defecto

Radial / Circular	1	2	3	4	5	6	7	10
A			/					
B								1
C								
D								
E	///		////. /					
F	/	//						9
G								3
H								
10	4	2	7					13

Figura 2.3 Hoja de registro de la localización de defectos

fectos fue 62 porque algunas veces se encontraron dos o más defectos en una misma pieza.

Ejemplo 2.3
Hoja de registro de localización de defectos

Los defectos externos, tales como los rayones y las manchas, se encuentran en toda clase de productos y en muchas plantas se están haciendo esfuerzos por reducir este tipo de defecto. La hoja de registro de la localización de defectos tiene un papel importante en la solución de este tipo de problema. Generalmente, muchas hojas de registro de este tipo tienen diagramas o ilustraciones ampliadas en las cuales se hacen los registros, de manera que pueda observarse el sitio de la ocurrencia del defecto.

La figura 2.3 muestra uno de estos ejemplos, utilizado por un fabricante de maquinaria en la inspección de partes metálicas, para aceptarlas o rechazarlas según sus características de porosidad. En el pasado se le informaba al proveedor únicamente

sobre el rechazo o la aceptación de cada lote y el número de defectos por lote, pero la calidad no mejoraba en absoluto.

Después de haber utilizado hojas de registro como la de la figura, que sirvieron como informes de inspección que indicaban dónde sería más probable que se presentara la porosidad, la calidad mejoró notablemente porque resultaba más fácil encontrar las causas de los defectos. Esta hoja de registro lleva fácilmente a la acción y es indispensable para el diagnóstico del proceso, porque frecuentemente es posible encontrar las causas de los defectos examinando las láminas en las cuales se presentan y observando cuidadosamente el proceso para determinar por qué los defectos se concentran en esos lugares.

Ejemplo 2.4
Hoja de registro de las causas del defecto

La hoja de registro del ejemplo anterior se usa para definir la localización de los defectos. Además, las hojas de registro se usan a veces para estratificación adicional con el fin de encontrar las causas de los defectos. Generalmente, la mayoría de los estudios cuyo objetivo es encontrar las causas de los defectos implican combinar los datos sobre las causas con los datos correspondientes de los efectos, manteniéndolas en un orden claramente correspondiente y analizándolas luego mediante la estratificación por causas o haciendo diagramas de dispersión. Sin embargo, este tipo de manejo de datos es posible hacerlo en una hoja de registro si el caso es sencillo.

Por ejemplo, la figura 2.4 es una hoja para registrar la ocurrencia de los defectos en ciertos picaportes, asignables a maquinaria, trabajadores, días y tipo de defecto. Podemos ver de inmediato que el trabajador B produce muchas unidades defectuosas. El miércoles, todos los trabajadores producen muchas unidades defectuosas. Una búsqueda de las causas mostró que el trabajador B no estaba cambiando los dados con suficiente frecuencia y que los miércoles las materias primas tenían una composición que tenía más probabilidad de causar defectos.



Equipo	Operario	Lunes		Martes		Miércoles		Jueves		Viernes		Sábado	
		AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM
Máquina 1	A	oo × •	o ×	ooo	o ××	ooo ××× •	oooo ×××	oooo × ••	o ××	oooo	oo	o	×× •
	B	o ×× •	ooo ××× •	oooooo ××	ooo ××	oooooo ×× •	oooooo × •	ooooo ××	ooo × ••	oo ×× •	ooooo	oo ×	oooo ×× •
Máquina 2	C	oo ×	o ×	oo	•	ooooo	oooooo ×	oo	o •	oo	oo △ □	o	o △ □
	D	oo ×	o ×	oo △	ooo •	ooo △ •	ooooo ×	oo •	oo △	oo △△ □	o ••	oo × □	o ××

o : Rayado en la superficie   × : Porosidad   △ : Terminado defectuoso
● : Forma inadecuada   □ : Otros

Figura 2.4 Hoja de registro de las causas del defecto

Con el mismo propósito, también es posible elaborar una hoja de registro usando el diagrama de causa-efecto. Es decir, puede prepararse un diagrama de causa-efecto que los trabajadores puedan comprender con facilidad y cuando se reconozca la causa de un defecto o la condición en la cual se produce, se hace una marca de registro en el espacio que se encuentra cerca de

cada flecha. A partir de ahí, es posible determinar cuál de las causas requiere acción prioritaria.

Además de los ejemplos descritos, hay muchas otras clases de hoja de registro que se utilizan en las fábricas. Las hojas de registro se diseñan considerando primero el propósito de la recolección de los datos y haciendo luego varias modificaciones creativas a fin de que los datos puedan recogerse y registrarse fácilmente y de manera adecuada al objetivo.

Ejercicio 2.1

En un proceso de pulimento de lentes hay dos trabajadores, y cada uno de ellos maneja dos máquinas. Recientemente, la fracción de productos defectuosos ha aumentado. Los trabajadores piden un cambio de maquinaria, diciendo que las máquinas actualmente utilizadas son demasiado viejas. Los gerentes encargados del proceso dicen que los trabajadores deben tener más cuidado porque están cometiendo errores por descuido. ¿Qué haría usted en esta situación?

III

El análisis de Pareto

3.1 ¿QUÉ SON LOS DIAGRAMAS DE PARETO?

Los problemas de calidad se presentan como pérdidas (productos defectuosos y su costo). Es muy importante aclarar el patrón de la distribución de la pérdida. La mayoría de las pérdidas se deberán a unos pocos tipos de defectos, y estos defectos pueden atribuirse a un número muy pequeño de causas. Si se identifican las causas de estos *pocos* defectos *vitales*, podremos eliminar casi todas las pérdidas, concentrándonos en esas causas particulares y dejando de lado por el momento otros *muchos* defectos *triviales*. El uso del diagrama de Pareto permite solucionar este tipo de problema con eficiencia.

En 1897, el economista italiano V. Pareto presentó una fórmula que mostraba que la distribución del ingreso es desigual. En 1907, el economista norteamericano M. C. Lorenz expresó una teoría similar por medio de diagramas. Estos dos estudiosos indicaron que una proporción muy grande del ingreso está en manos de muy pocas personas. Mientras tanto, en el campo del control de calidad, el Dr. J. M. Juran aplicó el método del diagrama de Lorenz como fórmula para clasificar los problemas de calidad en los *pocos vitales* y los *muchos triviales*, y llamó este méto-



do *análisis de Pareto*. Señaló que, en muchos casos, la mayoría de los defectos y de su costo se deben a un número relativamente pequeño de causas.

3.2 CÓMO ELABORAR DIAGRAMAS DE PARETO

Paso 1

Decida qué problemas se van a investigar y cómo recoger los datos.

1) Decida qué clase de problemas son los que usted quiere investigar.
   Ejemplo: Objetos defectuosos, pérdidas en términos monetarios, ocurrencia de accidentes.
2) Decida qué datos va a necesitar y cómo clasificarlos.
   Ejemplo: Por tipo de defecto, localización, proceso, máquina, trabajador, método.
   Nota: Resuma los ítems que se presentan con poca frecuencia en la categoría "otros".
3) Defina el método de recolección de los datos y el período de duración de la recolección.
   Nota: Se aconseja utilizar un formato de investigación.

Paso 2

Diseñe una tabla para conteo de datos, con espacio suficiente para registrar los totales (tabla 3.1).

Tipo de defecto	Conteo	Total
Fractura	卌 卌	10
Rayado	卌 卌 卌 卌 .... 卌 //	42
Mancha	卌 /	6
Tensión	卌 卌 卌 卌 .... 卌 ////	104
Rajadura	////	4
Burbuja	卌 卌 卌 卌	20
Otros	卌 卌 ////	14
Total		200

Tabla 3.1 Tabla para conteo de datos

Paso 3

Diligencie la tabla de conteo y calcule los totales.

Paso 4

Elabore una tabla de datos para el diagrama de Pareto con la lista de ítems, los totales individuales, los totales acumulados, la composición porcentual y los porcentajes acumulados (tabla 3.2).

Paso 5

Organice los ítems por orden de cantidad y llene la tabla de datos.

Nota: El ítem "otros" debe ubicarse en el último renglón, independientemente de su magnitud. Esto se debe a que está compuesto de un grupo de ítems, cada uno de los cuales es más pequeño que el menor de los ítems citados individualmente.

Paso 6

Dibuje dos ejes verticales y un eje horizontal.

1) Ejes verticales
   a) Eje izquierdo
      Marque este eje con una escala desde 0 hasta el total general.

Tipo de defecto	Número de defectos	Total acumulado	Composición porcentual	Porcentaje acumulado
Tensión	104	104	52	52
Rayado	42	146	21	73
Burbuja	20	166	10	83
Fractura	10	176	5	88
Mancha	6	182	3	91
Rajadura	4	186	2	93
Otros	14	200	7	100
Total	200	—	100	—

Tabla 3.2 Tabla de datos para un diagrama de Pareto

b) Eje derecho
Marque este eje con una escala desde 0% hasta 100%.

c) Eje horizontal
Divida este eje en un número de intervalos igual al número de ítems clasificados.

Paso 7

Construya un diagrama de barras.

Paso 8

Dibuje la curva acumulada (curva de Pareto).

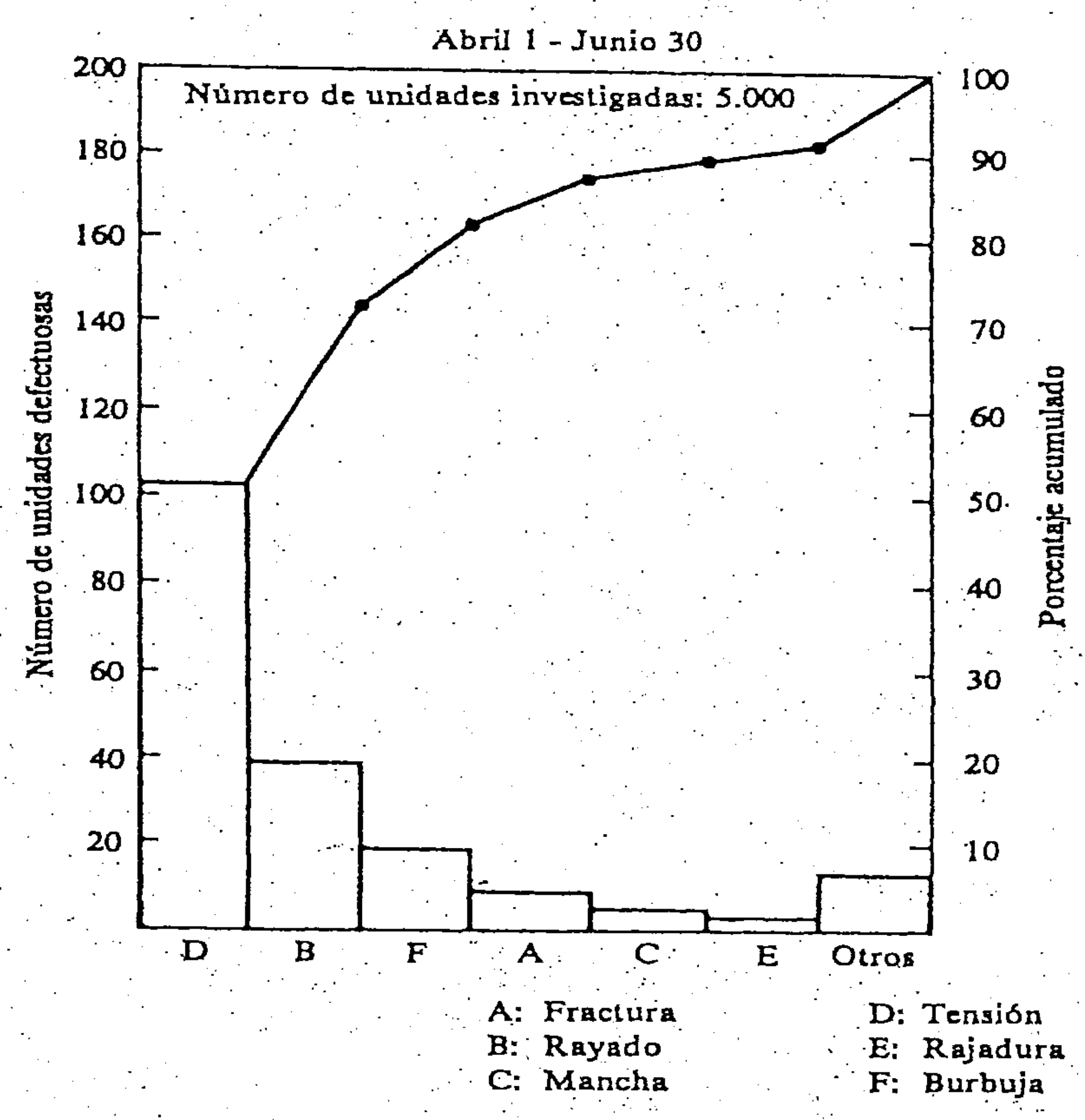


Figura 3.1 Diagrama de Pareto por ítems defectuosos

Marque los valores acumulados (total acumulado o porcentaje acumulado) en la parte superior, al lado derecho de los intervalos de cada ítem, y conecte los puntos con una línea continua.

Paso 9

Escriba en el diagrama cualquier información necesaria.

1) Información sobre el diagrama
   Título, cifras significativas, unidades, nombre del dibujante.
2) Información sobre los datos
   Período de tiempo, tema y lugar de la investigación, número total de datos.

3.3 DIAGRAMAS DE PARETO DE FENÓMENOS Y DIAGRAMAS DE PARETO DE CAUSAS

Como a se dijo, un diagrama de Pareto es un método para identificar los *pocos vitales*, y existen dos tipos.

(1) Diagramas de Pareto de fenómenos

Éste es un diagrama en el cual se relacionan los resultados indeseables, como los que se presentan a continuación, y se utiliza para averiguar cuál es el principal problema.

1) Calidad: Defectos, faltas, fracasos, quejas, ítems devueltos, reparaciones.
2) Costo: Magnitud de las pérdidas, gastos.
3) Entrega: Escasez de inventarios, demoras en los pagos, demoras en la entrega.
4) Seguridad: Accidentes, errores, interrupciones.

(2) Diagramas de Pareto de causas

Éste es un diagrama en el cual se relacionan los resultados indeseables, como los que se presentan a continuación, y se utiliza para averiguar cuál es el principal problema.



1) Operario: Turno, grupo, edad, experiencia, destreza.
2) Máquina: Máquinas, equipos, herramientas, organizaciones, modelos, instrumentos.
3) Materia prima: Productor, planta, lote, clase.
4) Método operacional: Condiciones, órdenes, disposiciones, métodos.

3.4 NOTAS SOBRE LOS DIAGRAMAS DE PARETO

(1) Sugerencias para elaborar diagramas de Pareto

1) Pruebe varias clasificaciones y construya muchas clases de diagramas de Pareto.
Usted podrá captar la esencia de un problema observándolo desde varios ángulos; es necesario tratar de encontrar varios métodos de clasificación hasta que identifique los *pocos vitales*, lo cual constituye el propósito del análisis de Pareto.
2) No es conveniente que "otros" represente un porcentaje de los más altos. Si esto sucede, se debe a que los ítems para la investigación no se han clasificado apropiadamente y demasiados ítems caen en esta categoría. En este caso, debe considerarse un método diferente de clasificación.
3) Si los datos se pueden representar en valores monetarios, lo mejor es dibujar diagramas de Pareto que muestren esto en el eje vertical. Si no se aprecian adecuadamente las implicaciones financieras de un problema, la investigación puede resultar ineficaz. En la administración, los costos constituyen una importante escala de medición.

(2) Sugerencias para usar diagramas de Pareto

1) Si un ítem se puede solucionar fácilmente, debe afrontarse de inmediato aunque sea relativamente de poca importancia. Debido a que un diagrama de Pareto tiene como objetivo la solución eficiente de problemas, se requiere, básicamente, que



afrontemos los *pocos vitales*. Sin embargo, si por medio de una sencilla medida se puede solucionar un ítem que parece relativamente de poca importancia, servirá como ejemplo de solución eficiente de un problema, y la experiencia, la información y los incentivos que los empleados pueden obtener por este medio serán de gran ayuda en la futura solución de problemas.

2) No deje de hacer un diagrama de Pareto de causas. Después de haber identificado el problema por medio de un diagrama de Pareto de fenómenos, para solucionarlo es necesario identificar las causas. Por tanto, es vital hacer un diagrama de Pareto de causas si es que se va a hacer mejoras.

Ejercicio 3.1

Analice la información de la tabla 3.3 haciendo varios diagramas de Pareto.

Operario	Máquina	Lunes	Martes	Miércoles	Jueves	Viernes
A	No. 1	•••• ** OO ## ☆	••••• * OOO #	••••• ***** OOOO ## ☆	•••• * OOO ##	••••• * OOOO ###
	No. 2	•• * O	••• ** OO ☆	••• ***** OO #	•• * O #	•• ** O #
B	No. 3	•• ** O #	•••• * O	••• ***** O #	••• * O # ☆	•••• * OO #
	No. 4	•• * OO ☆	••• * O #	••• **** OO #	••• * #	•• * OO #

• Tensión * Rayado O Burbuja # Fractura ☆ Otros

Tabla 3.3

IV

DIAGRAMAS DE CAUSA-EFECTO

4.1 ¿QUÉ SON LOS DIAGRAMAS DE CAUSA-EFECTO?

El resultado de un proceso puede atribuirse a una multitud de factores, y es posible encontrar la relación causa-efecto de esos factores. Podemos determinar la estructura o una relación múltiple de causa-efecto observándola sistemáticamente. Es difícil solucionar problemas complicados sin tener en cuenta esta estructura, la cual consta de una cadena de causas y efectos, y el método para expresar esto en forma sencilla y fácil es un diagrama de causa-efecto.

En 1953, Kaoru Ishikawa, profesor de la Universidad de Tokio, resumió la opinión de los ingenieros de una planta dándole la forma de un diagrama de causa-efecto mientras discutían un problema de calidad. Se dice que ésta fue la primera vez que se usó este enfoque. Antes de esto, el grupo de trabajo del profesor Ishikawa había usado este método para organizar los factores en sus actividades de investigación. Cuando el diagrama se usó en la práctica, mostró ser muy útil y pronto llegó a usarse ampliamente en muchas compañías en todo Japón. Se incluyó en la termi-



nología del JIS (Estándares Industriales Japoneses) del Control de Calidad, y se definió de la manera siguiente:

Diagrama de causa-efecto: Diagrama que muestra la relación entre una característica de calidad y los factores.

Actualmente, el diagrama se usa no solamente para observar las características de calidad de los productos sino también en otros campos, y ha sido ampliamente aplicado en todo el mundo.

4.2 CÓMO ELABORAR DIAGRAMAS DE CAUSA-EFECTO

Elaborar un diagrama de causa-efecto que sea útil no es tarea fácil. Puede afirmarse que quienes tienen éxito en la solución de problemas de control de calidad son aquéllos que tienen éxito en hacer diagramas de causa-efecto que sean útiles. Hay muchas maneras de hacer el diagrama, pero aquí se describirán dos métodos típicos.

Antes de presentar los procedimientos, se explicará la estructura del diagrama de causa-efecto con un ejemplo.

(1) Estructura y ejemplo de los diagramas de causa-efecto

Un diagrama de causa-efecto también se llama "Diagrama de espina de pescado", porque se parece al esqueleto de un pez, como se ve en la figura 4.1. Ocasionalmente se denomina también

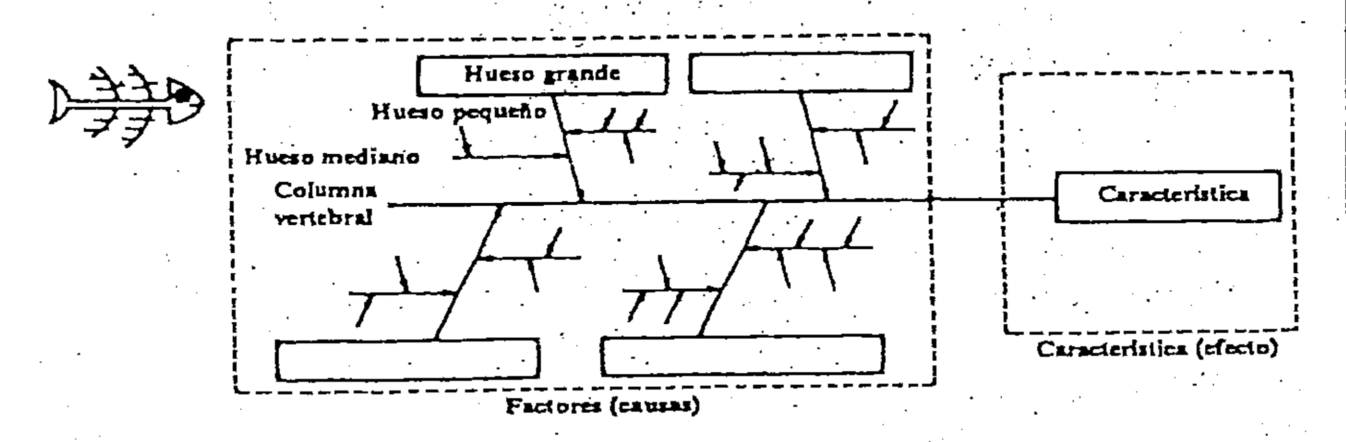


Figura 4.1 Estructura del diagrama de causa-efecto

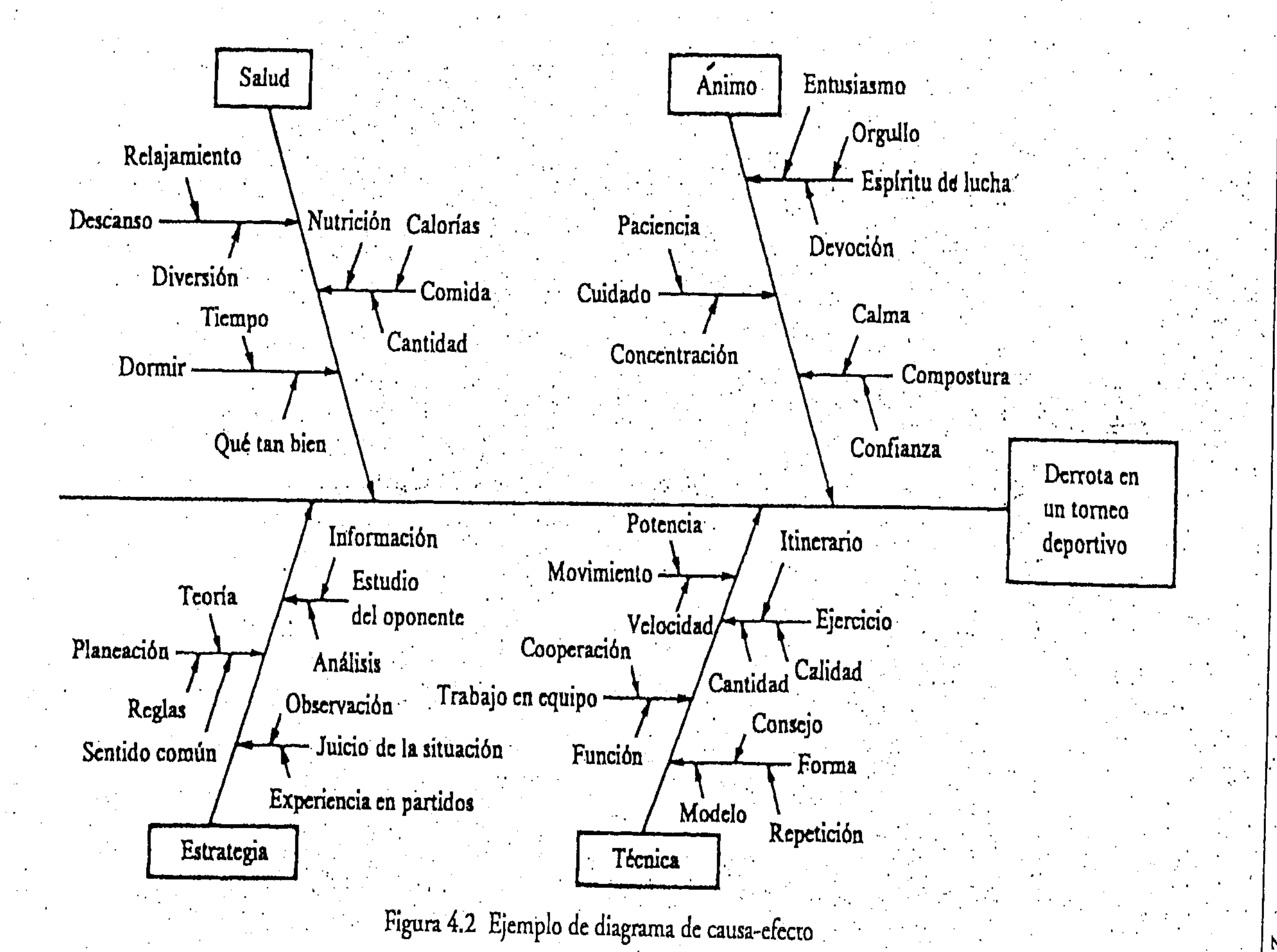


Figura 4.2 Ejemplo de diagrama de causa-efecto

diagrama de "árbol" o de "río", pero aquí se usa el nombre de "espina de pescado". En la figura 4.2 se muestra un ejemplo.

(2) Procedimiento para elaborar los diagramas de causa-efecto para la identificación de causas

Procedimiento

Paso 1
Describa el efecto o atributo de calidad.

Paso 2
Escoja una característica de calidad y escríbala en el lado derecho de una hoja de papel, dibuje de izquierda a derecha la línea de la espina dorsal y encierre la característica en un cuadrado. En seguida, escriba las causas primarias que afectan a la característica de calidad, en forma de grandes huesos, encerrados también en cuadrados.

Paso 3
Escriba las causas (causas secundarias) que afectan a los grandes huesos (causas primarias) como huesos medianos, y escriba las causas (causas terciarias) que afectan a los huesos medianos como huesos pequeños.

Paso 4
Asigne la importancia de cada factor, y marque los factores particularmente importantes que parecen tener un efecto significativo sobre la característica de calidad.

Paso 5
Registre cualquier información que pueda ser de utilidad.

Explicación del procedimiento

Con frecuencia puede parecer difícil proceder cuando se utiliza este enfoque. El mejor método en ese caso es considerar la "variación". Por ejemplo, observe la variación en la característica de calidad cuando reflexiona en los huesos grandes. Si los datos muestran que esa variación existe, observe por qué existe. Una variación en el efecto puede ser causada por una variación en los factores. Este tipo de reflexión puede ser muy eficaz.

Por ejemplo, cuando usted está elaborando un diagrama de causa-efecto relacionado con cierto defecto, puede descubrir que hay una variación en el número de defectos que ocurren en días diferentes de la semana. Si el defecto ocurre con más frecuencia los lunes que en cualquier otro día de la semana, usted puede reflexionar como sigue: "¿Por qué ocurrió el defecto?", "¿por qué ocurrió el defecto con mayor frecuencia los lunes que en cualquier otro día de la semana?" Esto lo hará buscar factores que hacen que el lunes sea diferente de los otros días, lo cual le permitirá descubrir finalmente la causa del defecto.

La adopción de este método de reflexión en cada etapa de la investigación de la relación que existe entre la característica y los huesos grandes, los huesos grandes y los huesos medianos, y los huesos medianos y los huesos pequeños, hace posible construir con bases racionales un diagrama útil de causa-efecto.

Una vez completo el diagrama de causa-efecto, el paso siguiente es asignar la importancia de cada factor. Todos los factores del diagrama no se relacionan necesariamente en forma estrecha con la característica. Marque esos factores que parecen tener un efecto particularmente significativo sobre la característica.

Finalmente, incluya cualquier información que pueda ser de utilidad en el diagrama, tal como el título, el nombre del producto, el proceso o grupo, la lista de participantes, la fecha, etc.



(3) Procedimiento de elaboración de diagramas de causa-efecto mediante listas sistemáticas de causas

Procedimiento

Paso 1
Escoja la característica de calidad.

Paso 2
Busque todas las causas posibles que puedan afectar a la característica de calidad.

Paso 3
Agrupe las causas por la afinidad que tengan entre sí y elabore un diagrama de causa-efecto conectando aquellos elementos que parecen tener un efecto significativo sobre la característica de calidad.

Paso 4
Asigne la importancia a cada factor, y señale los factores particularmente importantes que parecen tener un efecto significativo sobre la característica de calidad.

Paso 5
Escriba cualquier información que pueda ser de utilidad.

Explicación del procedimiento

Este enfoque se caracteriza por la relación que establece entre dos actividades diferentes: la percepción de tantas causas como sea posible y su agrupación sistemática.

Para la percepción de causas se requiere una discusión abierta y activa, y un método eficaz para dirigir una reunión con este propósito es la *tormenta de ideas,* inventada por A. F. Osborn en los Estados Unidos.

En la elaboración del diagrama de causa-efecto, las causas se deben agrupar sistemáticamente procediendo de los huesos pe-



queños a los huesos medianos, y después de los huesos medianos a los huesos grandes.

4.3 NOTAS SOBRE LOS DIAGRAMAS DE CAUSA-EFECTO

(1) Sugerencias para elaborar los diagramas de causa-efecto

1) *Identifique todos los factores relevantes mediante consulta y discusión entre muchas personas.* Los factores que influyen más fuertemente sobre la característica deben determinarse entre aquellos que aparecen en el diagrama. Si se deja por fuera un factor en la etapa de discusión inicial, antes de que se construya el diagrama, no aparecerá más tarde. En consecuencia, la discusión entre todas las personas involucradas es indispensable para preparar un diagrama completo que no tenga misiones.
2) *Exprese la característica tan concretamente como sea posible.* La característica que se expresa en términos abstractos dará como resultado un diagrama de causa-efecto basado en generalidades. Aunque ese tipo de diagramas no tenga errores básicos desde el punto de vista de las relaciones causa-efecto, no será muy útil para resolver problemas reales.
3) *Haga un diagrama para cada característica.* Los errores en el peso y en la longitud del mismo producto tendrán estructuras diferentes de causa-efecto, y deben analizarse en dos diagramas separados. El intento de incluir todo en un solo diagrama dará como resultado un diagrama inmanejable por ser demasiado grande y complicado, lo cual hará que la solución de los problemas sea muy difícil.
4) *Escoja una característica y unos factores medibles.* Una vez completo el diagrama de causa-efecto, es necesario captar la fuerza de la relación causa-efecto en forma objetiva utilizando datos. Con este fin, tanto la característica como los factores causales deben ser medibles. Cuando es imposible medirlos,

usted debe tratar de hacerlos medibles, o encontrar características sustitutas.

5) *Descubra factores sobre los que sea posible actuar.* Si usted ha identificado una causa sobre la cual es imposible actuar, el problema no se solucionará. Si se ha de mejorar, las causas deben subdividirse hasta el nivel en el cual sea posible actuar sobre ellas, o de lo contrario su identificación será un ejercicio sin sentido.

(2) Sugerencias para el uso de los diagramas de causa-efecto

1) *Asigne la importancia de cada factor objetivamente con base en datos.* El examen de los factores con base en su propia habilidad y experiencia es importante, pero es peligroso juzgar su importancia únicamente con base en las percepciones o impresiones subjetivas. La mayoría de los problemas que pueden solucionarse usando ese enfoque ya se habrían podido solucionar y, por lo tanto, la mayoría de los problemas restantes no pueden solucionarse usando ese enfoque. La asignación objetiva de la importancia a los factores usando datos es más científico y más lógico.

2) *Trate de mejorar continuamente el diagrama de causa-efecto mientras lo usa.* La utilización de un diagrama de causa-efecto le ayudará a identificar las partes que deben ser verificadas, omitidas o modificadas, así como a descubrir las partes que deben agregarse. Trate repetidamente de mejorar su diagrama; finalmente obtendrá un diagrama realmente útil. Esto le permitirá solucionar problemas y, al mismo tiempo, le ayudará a mejorar su habilidad y a incrementar su conocimiento técnico.

4.4 DIAGRAMAS DE PARETO Y DIAGRAMAS DE CAUSA-EFECTO

Para la solución de problemas deben usarse varios métodos con-

juntamente, y la combinación de un diagrama de Pareto con un diagrama de causa-efecto es especialmente útil. El siguiente es un ejemplo típico.

(1) Selección de problemas

A continuación se encuentra un ejemplo que ilustra la revisión de ítems defectuosos en un proceso de manufactura, utilizando el diagrama de Pareto. Cuando se clasificaron los datos recogidos durante dos meses, se encontró que los defectos de dimensión eran los más numerosos, pues constituían el 48% del total. Por tanto, tratamos de reducir el número de ítems defectuosos, poniendo énfasis en los defectos de dimensión.

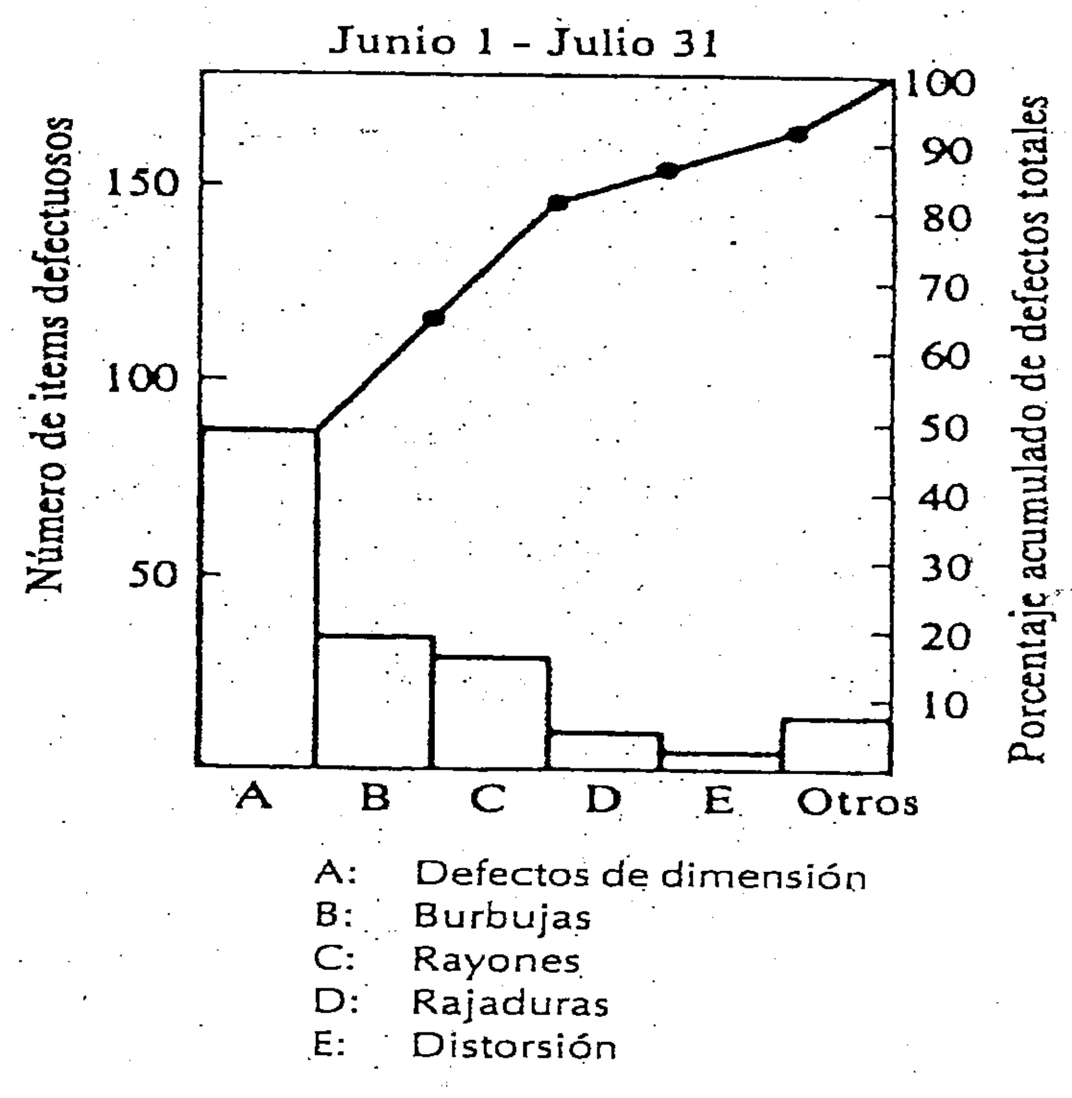


Figura 4.3 Diagrama de Pareto de ítems defectuosos

(2) Análisis y acciones correctivas

Todos los miembros del taller discutieron las causas de la variación dimensional y construyeron un diagrama de causa-efecto (ver figura 4.4). Después se hizo un diagrama de Pareto de causas (figura 4.5) investigando todas las unidades con variación dimensional para examinar hasta qué punto estos factores estaban

Variación dimensional
Máquina
Estabilidad
Operación
Mal balanceo
Deformación
Herramientas
Abrasión
Partes
Item
Inspección
Método
Método operacional
Posición
Material de ajuste
Angulo
Procedimiento
Trabajo
Velocidad
Grado de apretamiento
Ajuste
Orden
Operadores
Concentración
Espíritu
Atención al trabajo
Fatiga
Salud
Enfermedad
Entrenamiento
Educación
Habilidad
Experiencia
Partes y materiales
Figura
Forma
Diámetro
Dimensión
Componentes
Calidad de materiales
Almacenamiento

Figura 4.4 Diagrama de causa-efecto de los defectos de dimensión



afectando al defecto. Con algunos ítems era imposible aclarar las causas del defecto, y éstos se agruparon en la categoría "dudosos". Descubrimos, gracias al diagrama de Pareto, que la ocurrencia de los defectos se veía muy afectada por la posición de ajuste. Aunque la posición de ajuste se había estipulado en el estándar operacional tradicional, no se mostraba el método estándar de ajuste. Esto ocasionaba una variación en la posición de ajuste, y tenía como resultado los defectos dimensionales. Los miembros del taller diseñaron por tanto un método de ajuste apropiado, el cual se estandarizó y se incluyó en los estándares operacionales.

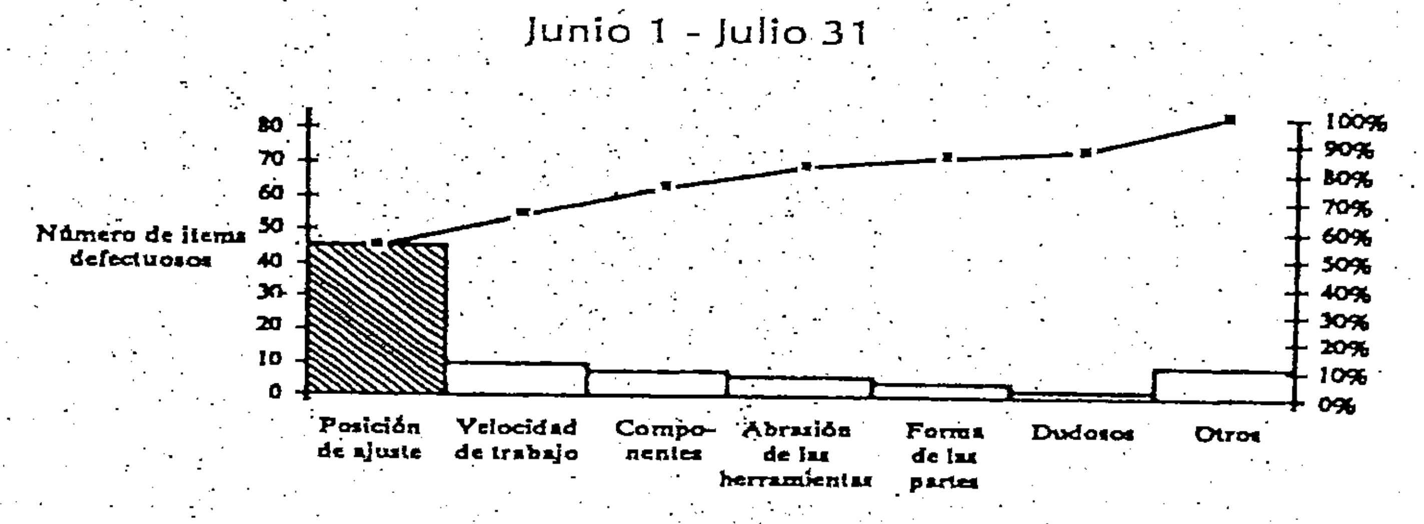


Figura 4.5 Diagrama de Pareto de causas

(3) Efectos de las mejoras

Después de haberse implementado los cambios, se recogieron datos y se hizo un diagrama de Pareto para comparar los resultados. Los dos diagramas de Pareto que aparecen en la figura 4.6 muestran claramente que los defectos dimensionales se redujeron.

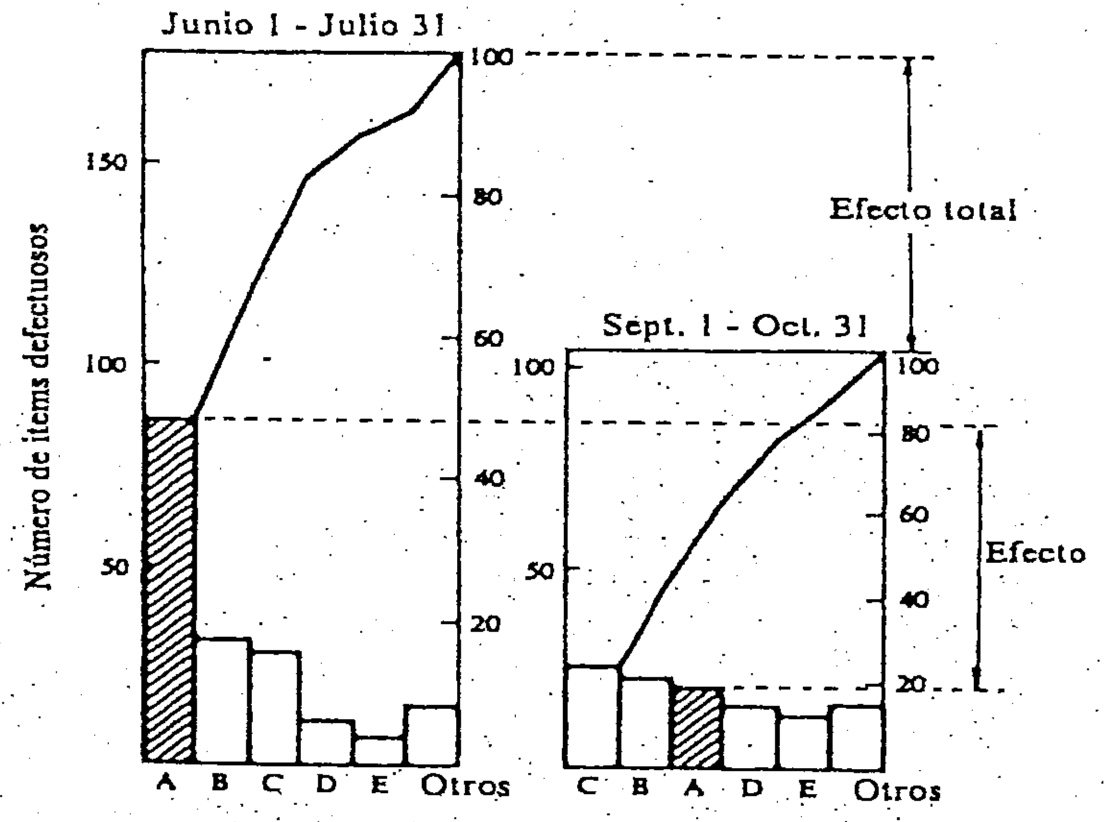


Figura 4.6 Comparación de diagramas de Pareto antes y después de las mejoras

Ejercicio 4.1

Haga un diagrama de causa-efecto de las siguientes características:

1) Errores de mecanografía.
2) Marcaciones equivocadas de un número telefónico.
3) Retraso para una cita.

V

Los histogramas

5.1 DISTRIBUCIONES E HISTOGRAMAS

(1) Variación y distribución

Si pudiéramos recoger datos sobre un proceso en el cual todos los factores (hombre, máquina, material, método, etc.) fueran perfectamente constantes, los datos sobre cada uno de estos factores conservarían su valor. Sin embargo, en la realidad es imposible mantener todos los factores constantes todo el tiempo. Estrictamente hablando, aun algunos factores que suponemos constantes no pueden ser perfectamente constantes. Es inevitable que los valores en un conjunto de información tengan variaciones. Los valores que toma un factor a través del tiempo no son siempre los mismos, pero eso no quiere decir que estén determinados de una manera desordenada. Aunque los valores cambian todo el tiempo, están gobernados por cierta regla, y ésta es que los datos tienen una determinada distribución.

(2) Población y muestras

En control de calidad, tratamos de descubrir los hechos reunien-



do datos y después tomamos las acciones apropiadas con base en esos hechos. Los datos no se recogen como un fin en sí mismos, sino como un medio para descubrir los hechos que están tras los datos.

Por ejemplo, consideremos el caso de una inspección por muestreo. Tomamos una muestra de un lote, realizamos un proceso de medición, y después decidimos si debemos aceptar todo el lote o no. En este caso, nuestra preocupación no es la muestra misma sino la calidad de todo el lote. Como otro ejemplo, consideremos el control de un proceso de manufactura utilizando una gráfica de control $\overline{x} - R$. Nuestro propósito no es determinar las características de la muestra con base en la cual hacemos la gráfica de control $\overline{x} - R$, sino averiguar cuál es el estado actual del proceso.

La totalidad de los ítems en consideración se denomina *población.* En el primer ejemplo anterior, la población es el lote, y en el segundo es el proceso.

A algunas personas puede parecerles difícil considerar un *proceso* como una *población* porque mientras que el lote es realmente un grupo finito de objetos individuales, un proceso no es de ninguna manera un producto, sino que se compone de cinco elementos (hombre, máquina, material, método y medición). Cuando fijemos nuestra atención en la función de fabricar productos, reconoceremos que sin duda el *proceso* produce un grupo de productos. Por otra parte, a menos que el proceso se detenga, el número de productos es infinito, razón por la cual se considera que un proceso es una *población* infinita.

Una *muestra* es uno o más ítems tomados de una población para proporcionar información sobre la población. Como una muestra se usa para estimar las características de toda la población, debe seleccionarse de tal manera que refleje las características de ésta. Un método común para la selección de muestras es seleccionar cualquier miembro de la población con igual probabilidad. Este método se llama *muestreo aleatorio,* y una muestra seleccionada por medio del muestreo aleatorio se denomina *muestra aleatoria.*



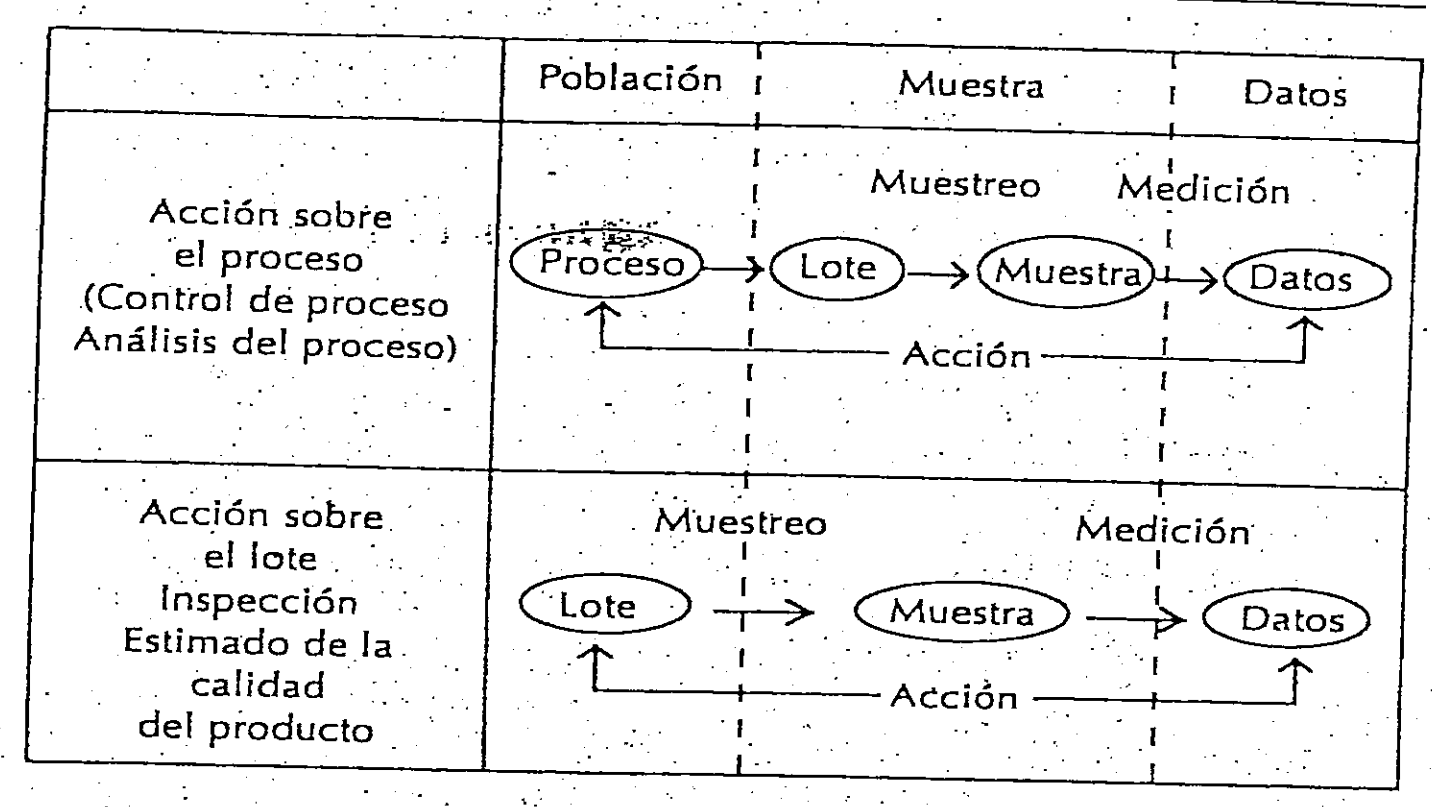


Figura 5.1 Población, muestra y datos

Los datos se obtienen midiendo las características de una muestra. Utilizando estos datos, llegamos a inferir sobre la población y, en consecuencia, tomamos una acción correctiva. Sin embargo, el valor obtenido de una muestra variará según la muestra seleccionada, lo cual dificulta decidir qué acción es apropiada. El análisis estadístico nos dirá cómo interpretar este tipo de datos. Los detalles se explicarán en el capítulo IX. La figura 5.1 muestra la relación entre población, muestra y datos.

(3) Histogramas

Los datos obtenidos de una muestra sirven como base para decidir sobre la población. Mientras más grande sea la muestra, más información obtendremos sobre la población. Pero un aumento en el tamaño de la muestra también implica un aumento en la cantidad de datos, y esto puede llegar a hacer difícil comprender la población a partir de esos datos, aun cuando se organicen en tablas. En ese caso, necesitamos un método que nos permita comprender la población de un vistazo. Un histograma responde a

esta necesidad. La organización de un buen número de datos en un histograma nos permite comprender la población de manera objetiva.

5.2 CÓMO ELABORAR HISTOGRAMAS

(1) Cómo construir tablas de frecuencia

Ejemplo 5.1

Para investigar la distribución de los diámetros de ejes de acero producidos en un proceso de laminación, se midieron los diámetros de 90 vigas, como se muestra en la tabla 5.1 Hagamos un histograma usando estos datos.

Muestra número	Resultados de la medición									
1 – 10	2.510	2.517	2.522	2.522	2.510	2.511	2.519	2.532	2.543	2.525
11 – 20	2.527	2.536	2.506	2.541	2.512	2.515	2.521	2.536	2.529	2.524
21 – 30	2.529	2.523	2.523	2.523	2.519	2.528	2.543	2.538	2.518	2.534
31 – 40	2.520	2.514	2.512	2.534	2.526	2.530	2.532	2.526	2.523	2.520
41 – 50	2.535	2.523	2.526	2.525	2.532	2.522	2.502	2.530	2.522	2.514
51 – 60	2.533	2.510	2.542	2.524	2.530	2.521	2.522	2.535	2.540	2.528
61 – 70	2.525	2.515	2.520	2.519	2.526	2.527	2.522	2.542	2.540	2.528
71 – 80	2.531	2.545	2.524	2.522	2.520	2.519	2.519	2.529	2.522	2.513
81 – 90	2.518	2.527	2.511	2.519	2.531	2.527	2.529	2.528	2.519	2.521

Tabla 5.1 Datos brutos

Procedimiento	Ejemplo
Paso 1 Calcule el rango (R) Obtenga el máximo y el mínimo de los valores observados y calcule R. R = (el máximo valor observado) – (el mínimo valor observado). El máximo y el mínimo de los valores observados se puede obtener fácilmente de la manera siguiente: Obtenga el máximo y el mínimo de los valores en cada una de las filas de la tabla de observaciones, y luego tome el mayor de los valores máximos y el menor de los valores mínimos. Éstos serán el máximo y el mínimo de todos los valores observados (tabla 5.2).	*Paso 1 Calcule R* R se obtiene a partir de los valores máximos y mínimos observados. (Ver tabla 5.2) Valor máximo = 2.545 Valor mínimo = 2.502 Por tanto, R = 2.545 – 2.502 = 0.043.

Paso 2 Determine el intervalo de clase El intervalo de clase se determina de manera que el rango, el cual incluye los valores máximo y mínimo, se divida en intervalos de igual amplitud. Para obtener la amplitud del intervalo, divida R por 1, 2 ó 5 (ó 10, 20, 50; 0.1, 0.2, 0.5, etc.),	*Paso 2 Determine el intervalo de clase* • 0.043 ÷ 0.002 = 21.5 y lo aproximamos al siguiente número entero, con lo cual tenemos 22. • 0.043 ÷ 0.005 = 8.6 y lo aproximamos al siguiente número entero, con lo cual tenemos 9.



Muestra número	Resultados de la medición										Valor máx. de la línea	Valor mín. de la línea
1 – 10	2.510	2.517	2.522	2.522	2.510	2.511	2.519	2.532	2.543	2.525	2.543	2.510
11 – 20	2.527	2.536	2.506	2.541	2.512	2.515	2.521	2.536	2.529	2.524	2.541	2.506
21 – 30	2.529	2.523	2.523	2.523	2.519	2.528	2.543	2.538	2.518	2.534	2.543	2.518
31 – 40	2.520	2.514	2.512	2.534	2.526	2.530	2.532	2.526	2.523	2.520	2.534	2.512
41 – 50	2.535	2.523	2.526	2.525	2.532	2.522	2.502	2.530	2.522	2.514	2.535	2.502
51 – 60	2.533	2.510	2.542	2.524	2.530	2.521	2.522	2.535	2.540	2.528	2.542	2.510
61 – 70	2.525	2.515	2.520	2.519	2.526	2.527	2.522	2.542	2.540	2.528	2.542	2.515
71 – 80	2.531	2.545	2.524	2.522	2.520	2.519	2.519	2.529	2.522	2.513	2.545	2.513
81 – 90	2.518	2.527	2.511	2.519	2.531	2.527	2.529	2.528	2.519	2.521	2.531	2.511
											El valor máx. 2.545	El valor mín. 2.502

Tabla 5.2 Tabla para el cálculo del rango



de manera que se obtengan entre 5 y 20 intervalos de clase de igual amplitud. Cuando haya dos posibilidades, use el intervalo de menor amplitud si el número de valores medidos es de 100 o más y el intervalo de mayor amplitud, si hay 99 o menos valores observados.

- 0.043 ÷ 0.010 = 4.3 y lo aproximamos al número entero más cercano, con lo cual tenemos 4.

De esta manera, el intervalo de clase se define como 0.005, pues esto da un número de intervalo entre 5 y 20. (Nota: 9 intervalos)*.

* Nota del editor en español.

Paso 3 Prepare el formato para la tabla de frecuencias

Prepare un formato como el de la tabla 5.3, en el cual se puedan registrar la clase, el punto medio, las marcas de frecuencia, la frecuencia, etc.

Paso 3 Prepare la tabla de frecuencias

Prepare una tabla como se muestra en la tabla 5.3.

Paso 4 Determine los límites de la clase

Determine los límites de los intervalos de manera que incluyan los valores mínimos y máximos, y escríbalos en la tabla de frecuencia. Primero, determine el límite inferior de la primera clase y súmele la amplitud del intervalo para obte-

Paso 4 Determine los límites de la clase

Los límites de la primera clase deben determinarse como 2.5005 y 2.5055 de manera que la clase incluya el valor mínimo 2.502; los límites de la segunda clase deben determinarse como 2.5055 – 2.5105, y así sucesivamente.

ner el límite entre la primera y la segunda clase. Cuando lo haga, cerciórese de que la primera clase contiene el valor mínimo. El límite inferior de la primera clase se ubica a 1/2 de la unidad de medida a partir del valor mínimo observado. Luego, siga sumando la amplitud del intervalo al valor previo para obtener el segundo límite, el tercero, y así sucesivamente, y cerciórese de que la última clase incluye el valor máximo.

Regístrelos en una tabla de frecuencias. (Ver tabla 5.3.)

(Nota: 2.502 – (0.005/2) = 2.4995 ~ 2.500, lo que daría un valor medio de la primera clase de 2.5025, que se redondea a 2.503 y se establecen los límites definitivos)*.

* Notas del editor en español.

Paso 5 Calcule el punto medio de la clase

Calcule el punto medio de la clase utilizando la siguiente ecuación, y escríbalo en la tabla de frecuencias.

Punto medio de la primera clase:

$$= \frac{\text{Suma de los límites superior e inferior de la primera clase}}{2},$$

Punto medio de la segunda clase:

$$= \frac{\text{Suma de los límites superior e inferior de la segunda clase}}{2},$$

Paso 5 Calcule el punto medio de la clase

Punto medio de la primera clase:

$$= \frac{2.5005 + 2.5055}{2} = 2.503,$$

Punto medio de la segunda clase:

$$= \frac{2.5005 + 2.5105}{2} = 2.508,$$

y así sucesivamente.

Los puntos medios de la segunda clase, la tercera clase, y así sucesivamente, pueden obtenerse también de la manera siguiente:

Punto medio de la segunda clase = punto medio de la primera + intervalo de clase. Punto medio de la tercera = punto medio de la segunda + intervalo de clase, y así sucesivamente.

Paso 6 Obtenga las frecuencias

Lea los valores observados uno por uno y registre las frecuencias correspondientes a cada clase, usando marcas en grupos de cinco, como sigue:

Frecuencia	1	2	3	4
Notación de frecuencia	/	//	///	////
Frecuencia	5	6	7	8
Notación de frecuencia	/	//	///	////

Paso 6 Obtenga las frecuencias

Registre las frecuencias (vea la tabla 5.3).



	Clase	Punto medio de la clase X	Marcas de frecuencia (conteo)	Frecuencia *f*
1	2.5005 – 2.5055	2.503	/	1
2	2.5055 – 2.5105	2.508	////	4
3	2.5105 – 2.5155	2.513	~~////~~ ////	9
4	2.5155 – 2.5205	2.518	~~////~~ ~~////~~ ////	14
5	2.5205 – 2.5255	2.523	~~////~~ ~~////~~ ~~////~~ ~~////~~ //	22
6	2.5255 – 2.5305	2.528	~~////~~ ~~////~~ ~~////~~ ////	19
7	2.5305 – 2.5355	2.533	~~////~~ ~~////~~	10
8	2.5355 – 2.5405	2.538	~~////~~	5
9	2.5405 – 2.5455	2.543	~~////~~ /	6
TOTAL			–	90

Tabla 5.3 Tabla de frecuencias

Notas

1. Habría un error en el conteo de las frecuencias si la suma de las frecuencias f ($\sum f$), no fuese igual al número total (n) de los valores observados.
2. Si se requiere la frecuencia relativa, puede obtenerse dividiendo la frecuencia f por n.

(2) Cómo elaborar un histograma

Procedimiento	Ejemplo *(Ejemplo 5.1)*

Paso 1

Sobre una hoja de papel cuadriculado, marque el eje horizontal con una escala. La escala no debe ser con base en el intervalo de clase; es mejor que sea con base en la unidad de



medición de los datos, por ejemplo, 10 gramos corresponderían a 10 milímetros.

Esto hace fácil la comparación con muchos histogramas que describan factores similares así como con las especificaciones (estándares). Deje un espacio aproximadamente igual al intervalo de clase en el eje horizontal a cada lado de la primera y de la última clase.

Paso 2

Marque el eje vertical de la izquierda con una escala de frecuencia y, si es necesario, dibuje el eje de la derecha y márquelo con una escala de frecuencias relativas. La altura de la clase con la frecuencia máxima debe ser entre 0.5 y 2.0 veces la distancia entre los valores máximo y mínimo en el eje horizontal.

Paso 3

Marque la escala horizontal con los límites de los valores de clase.

Paso 4

Utilizando los intervalos de clase como línea de base, dibuje un rectángulo cuya altura corresponda a la frecuencia en esa clase.

Paso 5

Dibuje una línea sobre el histograma para representar la media, y dibuje también una línea para representar el límite de especificación, si lo hay.

Paso 6

En un espacio en blanco del histograma, anote la historia de los datos (el período de tiempo durante el cual se recogieron los datos, etc.), el número de datos n, la media $\bar{x}$ y la desviación estándar s. (Ver figura 5.2.) Los cálculos para $\bar{x}$ y s se mostrarán en (3) de la sección 5.4.

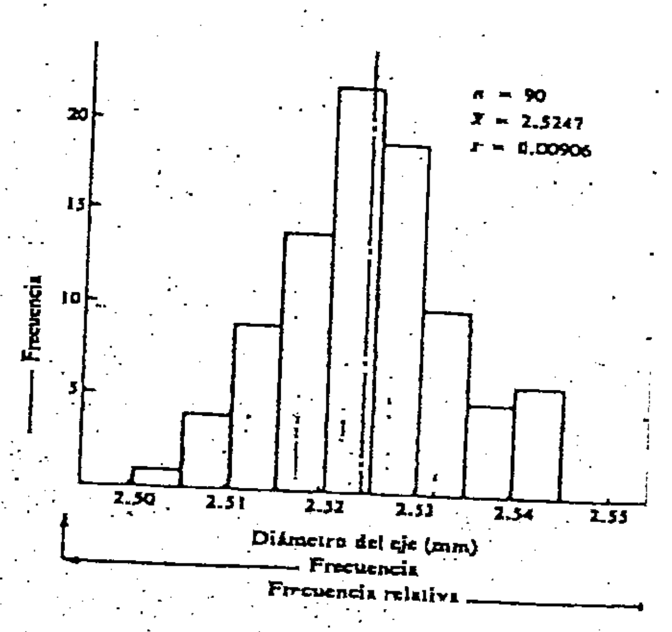


Figura 5.2 Histograma

5.3 CÓMO LEER HISTOGRAMAS

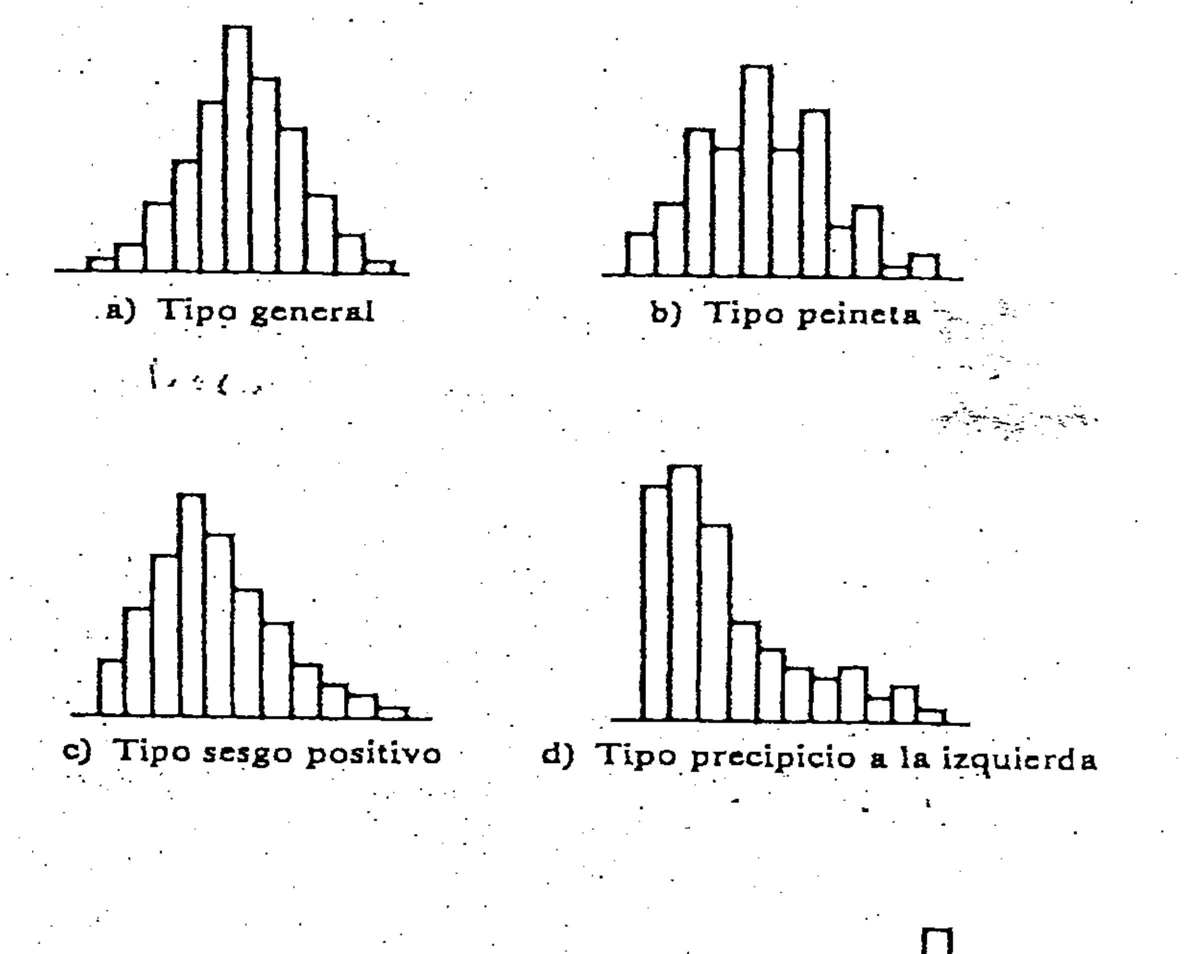
(1) Tipos de histogramas

Es posible obtener información útil sobre el estado de una población mirando la forma del histograma. Las siguientes son formas típicas, y podemos usarlas como indicios para analizar un proceso. (Ver figura 5.3.)

a) Tipo general (forma simétrica o de campana)

Forma: El valor de la media del histograma está en el centro de rango de los datos. La frecuencia es mayor en el centro y disminuye gradualmente hacia los extremos. La forma es simétrica.

Nota: Ésta es la forma más frecuente.



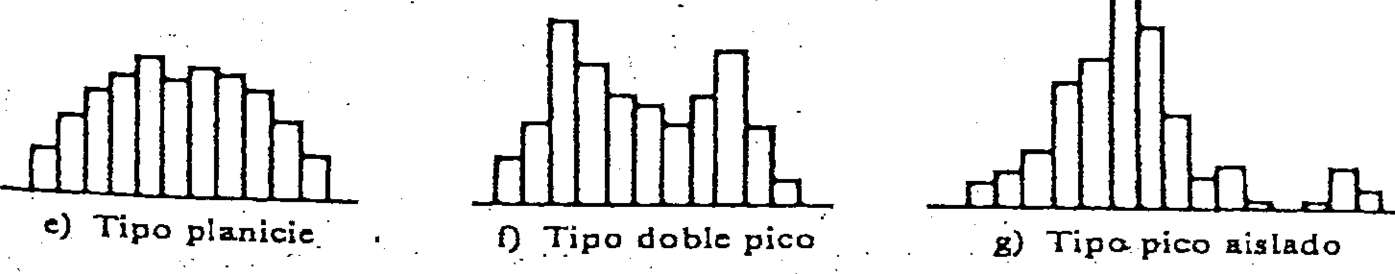


Figura 5.3 Tipos de histograma

b) Tipo peineta (multi-modal)

Forma: Cada tercera clase tiene una frecuencia menor.

Nota: Esta forma se presenta cuando el número de unidades de información incluida en la clase varía de una a otra o cuando hay una tendencia particular en la forma como se aproximan los datos.

c) Tipo con sesgo positivo (con sesgo negativo)

Forma: Asimétrica. El valor de la media del histograma está localizado a la izquierda (derecha) del centro del rango. La frecuencia disminuye de manera más bien brusca hacia la izquierda (derecha), pero gradualmente hacia la derecha (izquierda).

Nota: Esta forma se presenta cuando el límite inferior (superior) se controla teóricamente o por un valor de especificación o cuando no se presentan valores inferiores (superiores) a cierto valor.

d) Tipo de precipicio a la izquierda (de precipicio a la derecha)

Forma: Asimétrica. El valor de la media del histograma está localizado al extremo izquierdo (derecho), lejos del centro del rango. La frecuencia disminuye bruscamente a la izquierda (derecha), y gradualmente hacia la derecha (izquierda).

Nota: Ésta es una forma que se presenta frecuentemente cuando se ha realizado una selección de 100% debido a una baja capacidad del proceso, y también cuando el sesgo positivo (negativo) se hace aún más extremo.

e) Tipo planicie

Forma: Las frecuencias forman una planicie, porque las clases tienen más o menos la misma frecuencia, excepto aquéllas de los extremos.

Nota: Esta forma se presenta con una mezcla de varias distribuciones que tienen valores de la media diferentes.

f) Tipo de doble pico (bimodal)

Forma: La frecuencia es baja cerca del centro del rango de la información, y hay un pico a cada lado.

Nota: Esta forma se presenta cuando se mezclan dos distribuciones que tienen valores de la media muy diferentes.

g) Tipo de pico aislado

Forma: Se presenta un pequeño pico aislado además de un histograma de tipo general.

Nota: Ésta es la forma que se presenta cuando se incluye una pequeña cantidad de datos de una distribución diferente, como en el caso de anormalidad en el proceso, error de medición, o inclusión de información de un proceso diferente.

(2) La comparación de histogramas con los límites de especificación

Si hay una especificación, dibuje sobre el histograma, con líneas, los límites de la especificación, para comparar la distribución con la especificación. Luego observe si el histograma está localizado razonablemente dentro de los límites. Más adelante, en la figura 5.4, se describen cinco casos típicos. Utilice éstos como referencia para evaluar la población.

Cuando el histograma satisface la especificación,

a) Lo que se necesita es mantener el estado actual, puesto que el histograma satisface ampliamente la especificación.
b) Se satisface la especificación, pero no hay margen extra. Por tanto, es mejor reducir la variación en pequeño grado.

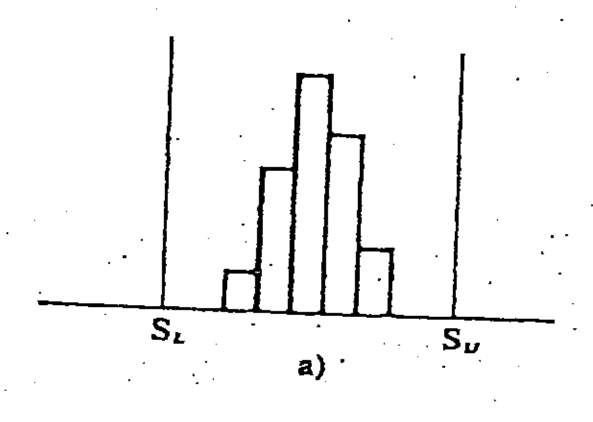
Cuando el histograma no satisface la especificación,

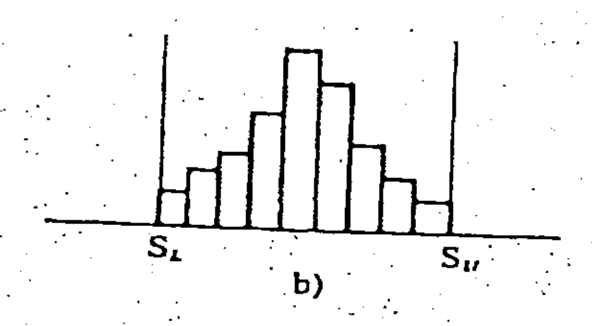
c) Es necesario tomar medidas para acercar la media al centro de la especificación.
d) Esto requiere de acciones para reducir la variación.
e) Se requieren las medidas descritas en c) y d) conjuntamente.

(3) Estratificación de histogramas

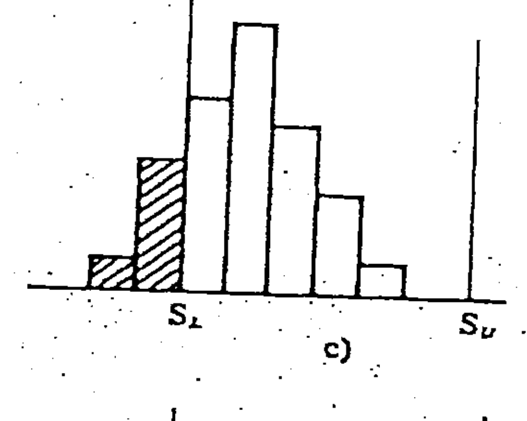
Cuando los valores observados se dividen en dos o más subpoblaciones, según la condición que existía en el momento de reco-

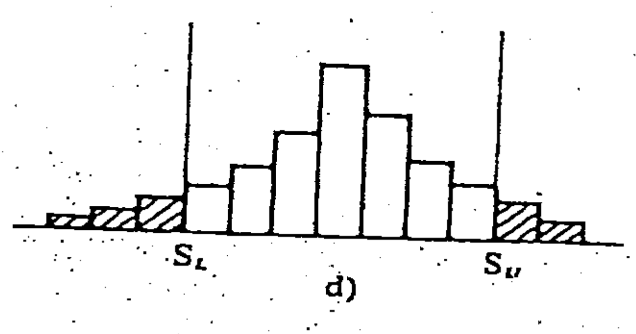
Casos en los cuales el histograma satisface la especificación





Casos en los cuales el histograma no satisface la especificación





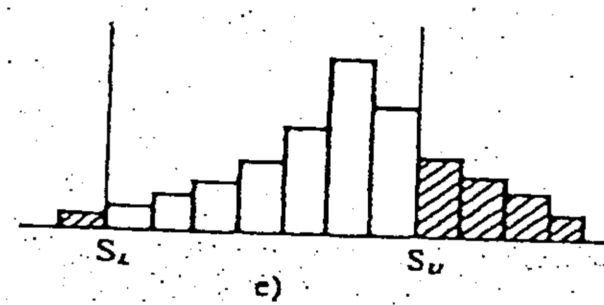


Figura 5.4 Histogramas y límites de la especificación

ger los datos, estas subpoblaciones se llaman *estratos,* y la división de los datos en estratos se llama *estratificación.*

Los valores observados siempre van acompañados de alguna variación. Por tanto, cuando los datos se estratifican según los factores que se cree pueden causar la variación, las causas de la variación se hacen más fácilmente detectables. Este método puede usarse efectivamente para mejorar la calidad del producto al reducir la variación y mejorar el promedio del producto.

Por lo general, la estratificación se hace según los materiales, las máquinas, las condiciones de operación y los trabajadores.

5.4 MEDIDAS PARA PRESENTAR LAS CARACTERÍSTICAS DE LAS DISTRIBUCIONES

(1) Medias y desviaciones estándar

El valor de las características medidas a una muestra tomada de una población variará, y no puede conocerse hasta que se observa. Este tipo de variable se llama *variable aleatoria*. Las características de calidad de los productos de las fábricas son de esa naturaleza.

Cuando se maneja ese tipo de datos, con frecuencia es más conveniente considerar los datos como un conjunto en lugar de tratar a cada ítem individualmente. Para poder ver los datos como un conjunto, determinamos primero cuál es el centro de los datos, y luego estudiamos cómo cada dato de la información se concentra alrededor del centro.

Una medida corriente para expresar el centro es la *media* o *expectativa*. Cuando hemos obtenido n datos, $x_1, \ldots, x_n$, la media de estos datos está dada por

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad (5.1)$$

y para todo el conjunto, la media está dada por

$$\mu = \sum x P(x), \qquad (5.2)$$

o

$$\mu = \int x f(x) dx, \qquad (5.3)$$

donde $P(x)$ es la probabilidad y $f(x)$ es la densidad probable de la variable aleatoria x.

$\bar{x}$ es la media de los datos recogidos, y se llama *media de la muestra*. μ es la media de la totalidad del conjunto que nos ocupa y se llama *media de la población*.

La *varianza* y la *desviación estándar* están entre las medias que se usan para expresar el grado de concentración de los datos

alrededor del centro. Cuando hemos obtenido n datos, $x_1, \ldots, x_n$, la varianza de estos datos se expresa por

$$V = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{5.4}$$

y la desviación estándar se expresa por

$$s = \sqrt{V}. \tag{5.5}$$

La varianza de una población está dada por

$$\sigma^2 = \sum (x - \mu)^2 P(x), \tag{5.6}$$

o

$$\sigma^2 = \int (x - \mu)^2 f(x) dx \tag{5.7}$$

y la desviación estándar, que es la raíz cuadrada de la varianza, se expresa como σ.

La varianza es la media del cuadrado de las diferencias entre el dato individual y la media. Una varianza grande significa que hay gran variación en los datos.

V y s son valores relacionados con los datos, y se llaman *varianza de la muestra* y *desviación estándar de la muestra*, respectivamente. σ^2 y σ son valores relacionados con una población y se llaman *varianza de la población* y *desviación estándar de la población*, respectivamente.

(2) El cálculo de medias y desviaciones estándar

Ejemplo 5.2

Las siguientes son las medidas de las dimensiones de cierta pieza de maquinaria. Calcule la media y la desviación estándar.

13.42 13.62 13.56 13.66 13.48 13.52 13.57

En algunos casos, puede realizarse una transformación de los datos

$$X_i = (x_i - a) \times h \tag{5.8}$$

para facilitar los cálculos. Entonces

$$\bar{x} = a + \frac{1}{h}\bar{X}, \tag{5.9}$$

$$S = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{h^2}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$= \frac{1}{h^2}\left\{\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right\}, \tag{5.10}$$

$$V = S/(n-1), \tag{5.11}$$

$$s = \sqrt{V}. \tag{5.12}$$

En este ejemplo, que a y h sean 13.40 y 100, respectivamente. Entonces obtenemos la siguiente tabla.

x	X	X^2
13.42	2	4
13.62	22	484
13.56	16	256
13.66	26	676
13.48	8	64
13.52	12	144
13.57	17	289
Total	103	1917

Tabla 5.4

$$\overline{X} = \frac{103}{7} = 14.7$$

De (5.9)

$$\bar{x} = 13.40 + \frac{1}{100} \times 14.7 = 13.547$$

De (5.10)

$$S = \frac{1}{100^2}\left\{1917 - \frac{1}{7} \times 103^2\right\} = 4.01 \times 10^{-2}$$

$$V = 4.01 \times 10^{-2} / (7-1) = 0.669 \times 10^{-2}$$

$$s = \sqrt{0.669 \times 10^{-2}} = 0.082$$

(3) El cálculo de las medias y las desviaciones estándar a partir de las tablas de frecuencia

Calculemos la media y la desviación estándar de los diámetros de 90 ejes como se muestra en la tabla 5.1. Cuando el número de datos es grande y éstos están resumidos en una tabla de frecuencia, el siguiente es el procedimiento para calcular la media y la desviación estándar.

Procedimiento	Ejemplo *(Ejemplo 5.1)*
Paso 1	
Prepare un formato de cálculo como en la tabla 5.5.	
Paso 2	
Escriba los límites de clase, los puntos medios de las clases y la frecuencia *f*.	



No.	Clase	Punto	Frecuencia f Medio x	u	uf	u^2f
1	2.5005 — 2.5055	2.503	1	− 4	− 4	16
2	2.5055 — 2.5105	2.508	4	− 3	− 12	36
3	2.5105 — 2.5155	2.513	9	− 2	− 18	36
4	2.5155 — 2.5205	2.518	14	− 1	− 14	14
5	2.5205 — 2.5255	2.523	22	0	0	0
6	2.5255 — 2.5305	2.528	19	1	19	19
7	2.5305 — 2.5355	2.533	10	2	20	40
8	2.5355 — 2.5405	2.538	5	3	15	45
9	2.5405 — 2.5455	2.543	6	4	24	96
Total			90	–	30	302

Tabla 5.5 Tabla de cálculo

Paso 3

Asigne el punto medio 0 ($u = 0$) a la clase que tiene la mayor f, y escriba 0 en la columna u.

Escriba −1, −2,... hacia los menores valores observados, y 1, 2... hacia los mayores valores observados.

La relación entre x y u se expresa en la siguiente ecuación:

$$u = \frac{(x - a)}{h} \qquad (5.13)$$

donde,

a: el punto medio de la clase donde $u = 0$

h: el intervalo de clase.

Paso 3

Asigne 0 al punto medio de u de la clase No. 5.

$a = 2.523$
$h = 0.005$



Paso 4

Registre los productos de u y de f en la columna uf, y los productos de u y de uf en la columna de u^2f; haga la suma de cada una, y regístrelas en los espacios correspondientes.

$$\sum uf = u_1f_1 + u_2f_2 + \ldots$$
$$\sum u^2f = u_1^2f_1 + u_2^2f_2 + \ldots$$

Paso 4

No. 1
$uf = (-4) \times 1 = -4$
No. 2
$uf = (-3) \times 1 = -12$

No. 1
$u^2f = uf \times u = (-4) \times (-4) = 16$
No. 2
$u^2f = uf \times u = (-12) \times (-3) = 36$

$\sum uf = (-4) + (-12) + \ldots + 24 = 30$
$\sum u^2f = 16 + 36 + \ldots + 96 = 302$

Paso 5

Calcule $\bar{x}$ usando la siguiente ecuación:

$$\bar{x} = a + h\,(\textstyle\sum uf/n). \qquad (5.14)$$

Paso 5

$$\bar{x} = 2.523 + 0.005 \times \frac{30}{90}$$
$$= 2.523 + 0.00167$$
$$= 2.52467 \text{ (mm)}$$

Paso 6

Calcule s usando la siguiente ecuación: (5.15)

$$s = h\sqrt{\left(\sum u^2f - \frac{(\sum uf)^2}{n}\right)/(n-1)}$$

Paso 6

$$s = 0.005 \times \sqrt{\left(302 - \frac{30^2}{90}\right)/(90-1)}$$
$$= 0.005 \times \sqrt{3.2809}$$
$$= 0.00906 \text{ (mm)}$$

5.5 LA DISTRIBUCIÓN NORMAL Y SUS CARACTERÍSTICAS

(1) La distribución normal

Un histograma se construye a partir de un cierto número de datos. Pero, ¿qué le pasaría al histograma si continuamos aumentando el número de datos? Si el intervalo de clase se reduce poco a poco a medida que se aumenta el número de datos, se obtiene una distribución de frecuencia lisa como límite de una distribución de frecuencia relativa. En realidad es una expresión de la población misma, puesto que se obtiene de un número infinito de datos.

Existen muchas clases de distribución, y la más típica es la *distribución normal*. En muchos casos, cuando la variación de una característica de calidad es causada por la suma de un gran número de errores infinitesimales independientes debidos a diferentes factores, la distribución de la característica de calidad se aproxima a una distribución normal. La forma de la distribución normal puede describirse sencillamente como una forma de campana o de montaña, y en una descripción más detallada,

a) La frecuencia es mayor en el centro y disminuye gradualmente hacia los extremos y,
b) es simétrica.

Esta curva puede expresarse matemáticamente como sigue:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (5.16)$$

La figura 5.5 muestra la forma de esta distribución.

(2) Características de la distribución normal

Como podemos ver en la ecuación (5.16), la ecuación de la distribución normal tiene dos parámetros: μ y σ^2.

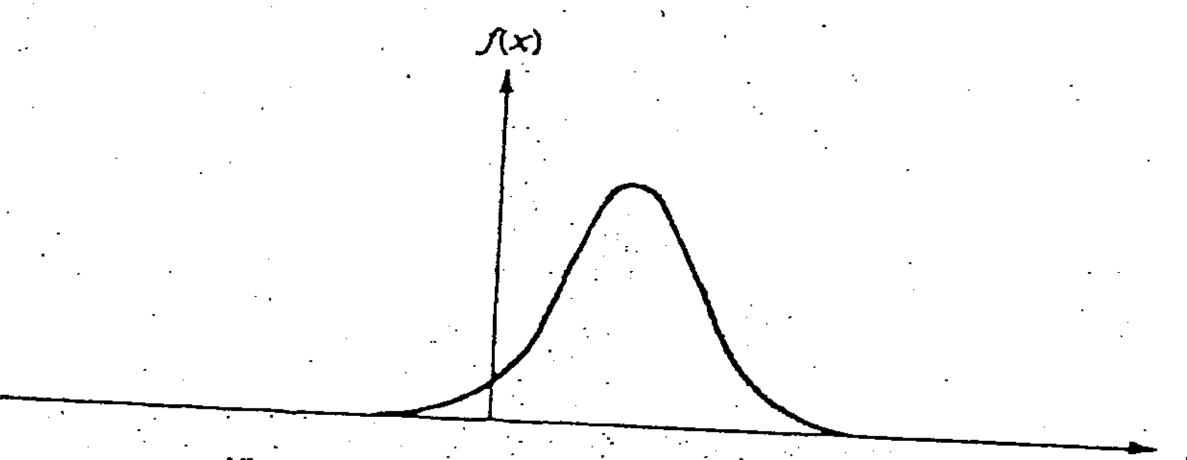


Figura 5.5 Forma de la distribución normal

La distribución normal se determina únicamente por estos dos parámetros y se designa simplemente por N (μ, σ^2). Estos dos parámetros tienen el siguiente significado:

μ: el centro de la distribución (la media)
σ: la dispersión de la distribución (la desviación estándar)

Pueden describirse gráficamente como en la figura 5.6.

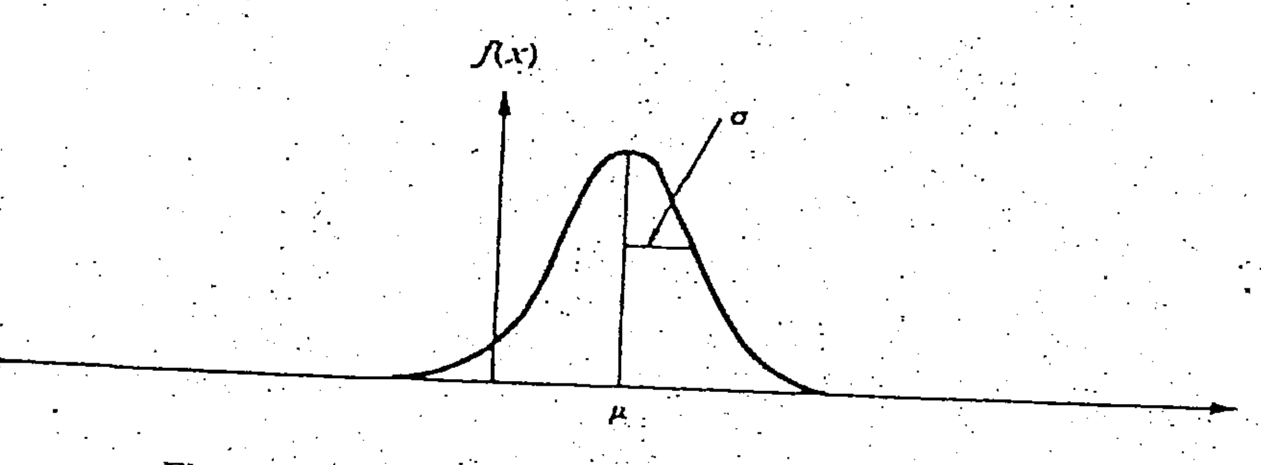


Figura 5.6 La distribución normal y sus parámetros

Para calcular una probabilidad en una distribución normal, estandarizamos y usamos la tabla de la distribución normal. *Estandarizar* es transformar una variable x a

$$u = \frac{(x - \mu)}{\sigma}. \qquad (5.17)$$

Tenemos entonces la medida estándar u, la cual se distribuye como la *distribución estándar normal* $N(0,1^2)$. La *tabla de la distribución normal* da las probabilidades en la distribución estándar normal. (Vea la tabla A. 1 en el apéndice de este libro.)

Consideremos la probabilidad de que una variable aleatoria x de $N(\mu, \sigma^2)$ caiga dentro de los límites $\mu \pm u\sigma$. La figura 5.7 muestra la probabilidad para varios valores de u. Teóricamente, una variable normal puede tener cualquier valor entre $-\infty$ y $+\infty$. Pero, por la figura tenemos 99.7% para $u = 3$. Esto quiere decir que en la práctica podemos despreciar la posibilidad de que x caiga fuera de los límites de $\mu \pm 3\sigma$. Este hecho es una regla importante de la distribución normal, y se llama la *regla de 3-sigma*. Es la base para determinar los límites de control en una hoja de control.

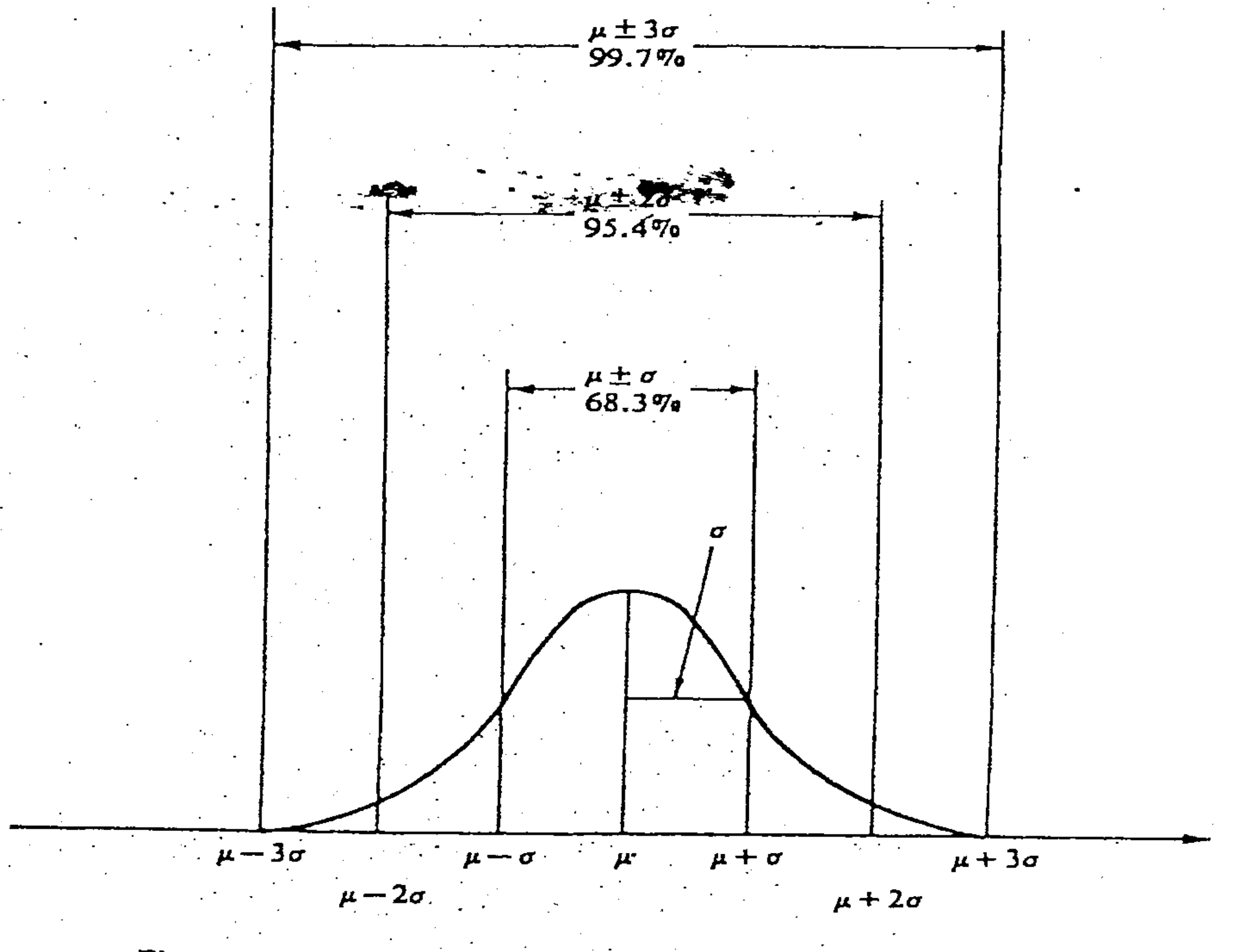


Figura 5.7 μ, σ y la probabilidad de la distribución normal

(3) Índice de capacidad del proceso

Con frecuencia, después de que el histograma muestra que sigue una distribución normal, se inicia un estudio de la capacidad del proceso. Esto se hace para saber si el proceso puede cumplir o no las especificaciones. Si suponemos que el proceso está distribuido normalmente, podemos de inmediato determinar el porcentaje de productos defectuosos a partir de las especificaciones dadas y de los parámetros (μ, σ), pero es más útil evaluar el proceso usando *Cp* (Índice de capacidad del proceso). La definición del *Cp* es la siguiente:

Especificaciones bilaterales (LE_s y LE_i)

$$Cp = \frac{LE_s - LE_i}{6s} \tag{5.18}$$

Especificaciones unilaterales (LE_s o LE_i)

$$Cp = \frac{LE_s - \bar{x}}{3s} \tag{5.19}$$

o

$$Cp = \frac{\bar{x} - LE_i}{3s} \tag{5.20}$$

Y la evaluación del proceso usando *Cp* es como sigue:

1)	$1.33 \leq Cp$	satisfactorio
2)	$1.00 \leq Cp < 1.33$	adecuado
3)	$Cp < 1.00$	inadecuado

Ejemplo 5.3

La tabla 5.6 muestra el producto de un cierto proceso de reacción química. Puesto que se usaron dos reactores, *A* y *B*, se señaló que tal vez había una diferencia entre ellos. Se estratificó según los reactores, y los resultados se muestran en la figura 5.8. Se encontraron diferencias entre los dos reactores.

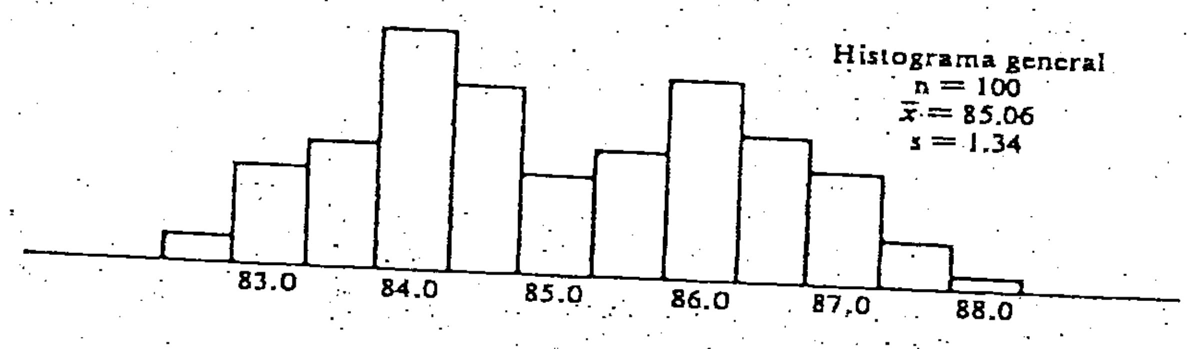


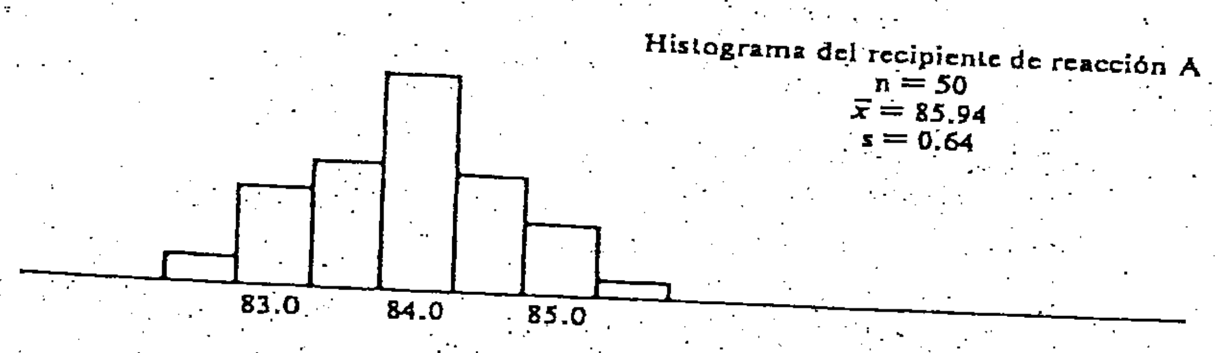
No.	R.V.	x	No.	R.V.	x	No.	R.V.	x	No.	R.V.	x
1	*A*	84.9	26	*B*	86.2	51	*B*	86.6	76	*B*	85.4
2	*A*	83.8	27	*B*	87.2	52	*B*	87.0	77	*B*	84.6
3	*B*	86.2	28	*A*	83.0	53	*B*	86.7	78	*A*	83.9
4	*B*	85.7	29	*B*	86.3	54	*A*	84.9	79	*A*	83.2
5	*A*	83.9	30	*A*	83.9	55	*A*	83.7	80	*B*	85.7
6	*B*	86.4	31	*A*	83.5	56	*B*	84.7	81	*B*	86.9
7	*B*	86.8	32	*B*	84.1	57	*A*	85.1	82	*A*	84.0
8	*B*	87.0	33	*B*	84.7	58	*B*	85.4	83	*B*	85.7
9	*A*	83.8	34	*A*	85.3	59	*A*	84.4	84	*A*	84.3
10	*B*	86.0	35	*A*	84.5	60	*A*	84.2	85	*B*	86.0
11	*B*	86.3	36	*A*	84.5	61	*B*	85.8	86	*A*	83.6
12	*A*	83.0	37	*B*	86.2	62	*A*	85.1	87	*B*	86.0
13	*A*	83.5	38	*A*	84.1	63	*A*	84.4	88	*A*	83.6
14	*A*	82.7	39	*A*	83.2	64	*A*	83.8	89	*B*	86.5
15	*B*	85.2	40	*B*	86.2	65	*B*	87.0	90	*B*	87.6
16	*B*	86.7	41	*A*	82.9	66	*B*	86.9	91	*A*	84.7
17	*A*	83.1	42	*A*	83.8	67	*B*	85.5	92	*A*	85.1
18	*B*	85.9	43	*A*	83.7	68	*A*	83.7	93	*A*	83.8
19	*B*	87.5	44	*B*	86.6	69	*B*	86.0	94	*B*	86.6
20	*A*	83.8	45	*B*	85.7	70	*A*	84.5	95	*B*	86.7
21	*B*	87.5	46	*A*	82.9	71	*B*	87.9	96	*A*	84.3
22	*A*	84.4	47	*B*	86.9	72	*A*	82.7	97	*A*	83.7
23	*A*	83.4	48	*B*	86.1	73	*A*	84.2	98	*B*	84.9
24	*A*	84.3	49	*B*	86.0	74	*A*	83.9	99	*B*	85.8
25	*B*	86.1	50	*A*	83.8	75	*B*	85.5	100	*B*	84.1

Tabla 5.6 Estratificación de la información

Ejercicio 5.1

En una panadería, dos trabajadores, *A* y *B*, están haciendo pan en las máquinas 1 y 2. El peso de los panes franceses producidos se registró durante 20 días, como se muestra en la tabla 5.7. Cada día se tomaron aleatoriamente cuatro panes de cada máquina y se pesaron. La especificación del peso es de 200-225 (g).





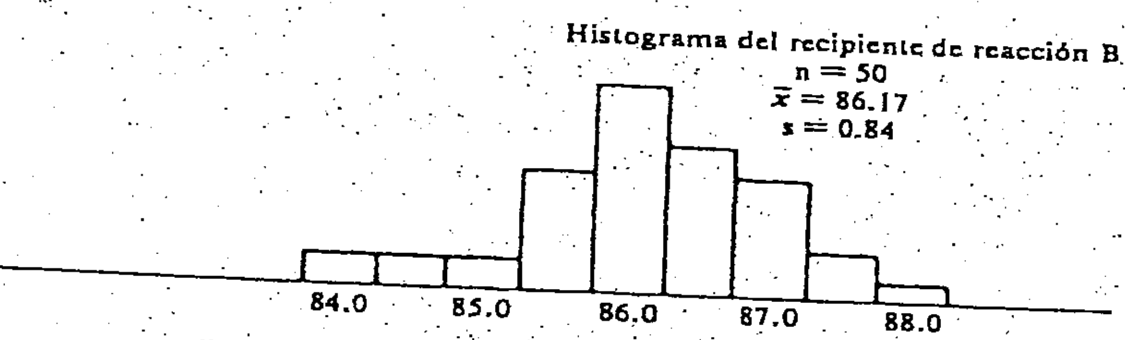


Figura 5.8 Estratificación de histogramas

1) Haga los siguientes histogramas:
   a) Un histograma global.
   b) Un histograma del panadero *A* y otro del panadero *B*.
   c) Un histograma de la máquina 1 y otro de la máquina 2.
   d) Cuatro histogramas que muestren combinaciones de diferentes panaderos y de diferentes máquinas.
2) Estúdielos comparándolos con la especificación.

Día	Panadero	Máquina No. 1				Máquina No. 2			
1	*A*	209.2	209.5	210.2	212.0	214.3	221.8	214.6	214.4
2	*A*	208.5	208.7	206.2	207.8	215.3	216.7	212.3	212.0
3	*A*	204.2	210.2	210.5	205.9	215.7	213.8	215.2	202.7
4	*B*	204.0	203.3	198.2	199.9	212.5	210.2	211.3	210.4
5	*B*	209.6	203.7	213.2	209.6	208.4	214.9	212.8	214.8
6	*A*	208.1	207.9	211.0	206.2	212.3	216.2	208.4	210.8
7	*A*	205.2	204.8	198.7	205.8	208.1	211.9	212.9	209.0
8	*B*	199.0	197.7	202.0	213.1	207.5	209.9	210.6	212.3
9	*B*	197.2	210.6	199.5	215.3	206.9	207.1	213.6	212.2
10	*B*	199.1	207.2	200.8	201.2	209.6	209.5	206.8	214.2
11	*A*	204.6	207.0	200.8	204.6	212.2	209.8	207.6	212.6
12	*B*	214.7	207.5	205.8	200.9	211.4	211.2	214.4	212.6
13	*B*	204.1	196.6	204.6	199.4	209.6	209.2	206.1	207.1
14	*A*	200.2	205.5	208.0	202.7	203.5	206.9	210.6	212.3
15	*A*	201.1	209.2	205.5	200.0	209.1	206.3	209.8	211.4
16	*A*	201.3	203.1	196.3	205.5	208.0	207.9	205.3	203.6
17	*B*	202.2	204.4	202.1	206.6	210.0	209.4	209.1	207.0
18	*B*	194.1	211.0	208.4	202.6	215.6	211.8	205.4	209.0
19	*B*	204.8	201.3	208.4	212.3	214.5	207.5	212.9	204.3
20	*A*	200.6	202.3	204.3	201.4	209.1	205.8	212.0	204.2

Tabla 5.7

VI

Los diagramas de dispersión

6.1 ¿QUÉ SON LOS DIAGRAMAS DE DISPERSIÓN?

En la práctica, frecuentemente es necesario estudiar la relación de correspondencia de dos variables. Por ejemplo, ¿hasta qué punto se afectarán las dimensiones de una parte de una máquina por el cambio en la velocidad de un piñón? O suponga que a usted le gustaría controlar la concentración de un material donde es preferible sustituir la medición de la concentración por la gravedad específica, porque prácticamente resulta más fácil medirla. Para estudiar la relación entre dos variables tales como la velocidad del piñón y las dimensiones de una parte, o la concentración y la gravedad específica, puede usarse lo que se llama un *diagrama de dispersión*.

Las dos variables que trataremos pueden enmarcarse así:

a) una característica de calidad y un factor que la afecta,
b) dos características de calidad relacionadas, o
c) dos factores relacionados con una sola característica de calidad.



Para comprender la relación entre éstas, es importante, en primer lugar, hacer un diagrama de dispersión y comprender la relación global.

6.2 CÓMO ELABORAR UN DIAGRAMA DE DISPERSIÓN

Para elaborar un diagrama de dispersión, se siguen los pasos siguientes:

Paso 1

Reúna pares de datos (x, y), cuyas relaciones usted quiere estudiar, y organice esa información en una tabla. Es deseable tener al menos 30 pares de datos.

Paso 2

Encuentre los valores mínimo y máximo para x y y. Decida las escalas que va a usar en los ejes horizontal y vertical de manera que ambas longitudes sean aproximadamente iguales, lo cual hará que el diagrama sea más fácil de leer. Trate de mantener el número de divisiones en cada eje entre 3 y 10 y use números redondos para facilitar la lectura. Cuando las dos variables sean un factor y una característica de calidad, use el eje horizontal x para el factor y el eje vertical y para la característica de calidad.

Paso 3

Registre los datos en el gráfico. Cuando se obtengan los mismos valores en diferentes observaciones, muestre estos puntos haciendo círculos concéntricos (◎), o registre el segundo punto muy cerca del primero.

Paso 4

Registre todos los aspectos que puedan ser de utilidad. Cerciórese de que se incluyan todos los ítems siguientes de manera que

cualquier persona, además de la persona que hizo el diagrama, pueda comprenderlo de un vistazo:

a) título del diagrama
b) periodo de tiempo
c) número de pares de datos
d) título y unidades de cada eje
e) nombre (etc.) de la persona que hizo el diagrama.

Ejemplo 6.1

Un fabricante de tanques de plástico que los hacía usando el método de moldeo por soplado tuvo dificultades con tanques defectuosos que tenían paredes delgadas. Se sospechaba que la causa de las paredes defectuosas era la variación en la presión del aire de soplado, la cual variaba cada día. La tabla 6.1 muestra la información sobre la presión del aire de soplado y el porcentaje de defectos. Hagamos un diagrama de dispersión con estos datos, siguiendo los pasos indicados arriba.

Paso 1

Como se ve en la tabla 6.1, hay 30 pares de datos.

Paso 2

En este ejemplo, sea x (eje horizontal) la presión del aire, y y (eje vertical) el porcentaje de defectos. Entonces:

el valor máximo de x: $x_{máx} = 9.4$ (kgf/cm^2),
el valor mínimo de x: $x_{mín} = 8.2$ (kgf/cm^2),
el valor máximo de y: $y_{máx} = 0.928$ (%),
el valor mínimo de y: $y_{mín} = 0.864$ (%)



Marcamos:

el eje horizontal en intervalos de 0.5 (kgf/cm²), desde 8.0 hasta 9.5 (kgf/cm²),

y

el eje vertical en intervalos de 0.01 (%), desde 0.85 hasta 0.93 (%).

Paso 3

Registre los datos (ver la figura 6.1):

Paso 4

Indique el intervalo de tiempo de la muestra obtenida (Oct.1 - Nov. 9), el número de muestras (n = 30), el eje horizontal (presión del aire de soplado [kgf/cm²]), el eje vertical (porcentaje de defectos [%]), y el título de diagrama (Diagrama de dispersión de presión del aire de soplado y porcentaje de defectos).

Fecha	Presión de aire (Kg/cm²)	Porcentaje de defectos (%)	Fecha	Presión de aire (Kg/cm²)	Porcentaje de defectos (%)
Oct. 1	8.6	0.889	Oct. 22	8.7	0.892
2	8.9	0.884	23	8.5	0.877
3	8.8	0.874	24	9.2	0.885
4	8.8	0.891	25	8.5	0.866
5	8.4	0.874	26	8.3	0.896
8	8.7	0.886	29	8.7	0.896
9	9.2	0.911	30	9.3	0.928
10	8.6	0.912	31	8.9	0.886
11	9.2	0.895	Nov. 1	8.9	0.908
12	8.7	0.896	2	8.3	0.881
15	8.4	0.894	5	8.7	0.882
16	8.2	0.864	6	8.9	0.904
17	9.2	0.922	7	8.7	0.912
18	8.7	0.909	8	9.1	0.925
19	9.4	0.905	9	8.7	0.872

Tabla 6.1 Datos de presión del aire de soplado y porcentaje de defectos de tanque plástico.



(%) Oct. 1 - Nov. 9
n = 30

0.93
0.92
0.91
0.90
0.89
0.88
0.87
0.86
0.85

% Defectos

8.0 8.5 9.0 9.5 (kgf/cm²)

Presión de aire (Kg/cm²)

Figura 6.1 Diagrama de dispersión de presión del aire de soplado y porcentaje de defectos.

6.3 CÓMO LEER LOS DIAGRAMAS DE DISPERSION

Así como es posible captar la forma de distribución en un histograma, también es posible leer la distribución general de los pares de datos a partir de un diagrama de dispersión. Al hacerlo, lo primero es examinar si hay o no hay puntos muy apartados en el diagrama. Puede generalmente suponerse que estos puntos apartados del grupo principal (figura 6.2) son el resultado de errores de medición o de registro de los datos, o fueron causados por algún cambio en las condiciones de operación. Es necesario excluir estos puntos del análisis correlacional. Sin embargo, en lugar de despreciar completamente estos puntos, usted debe dar la

debida atención a la causa de esas irregularidades porque con frecuencia se obtiene información muy útil averiguando por qué ocurren.

Hay muchos tipos de formas de dispersión, y algunas formas típicas se dan en la figura 6.3.1 hasta la 6.3.6. En las figuras 6.3.1 y 6.3.2, *y* crece con *x*; esto es una *correlación positiva*. También, como la figura 6.3.1 muestra esta tendencia de una manera muy pronunciada, se dice que es una correlación positiva fuerte. Las figuras 6.3.4 y 6.3.5 muestran lo opuesto a una correlación positiva, pues a medida que *x* aumenta, *y* disminuye; esto se llama una *correlación negativa*. La figura 6.3.4 muestra una correlación negativa fuerte. La figura 6.3.3 muestra el caso en el que *x* y *y* no tienen ninguna relación particular, y por eso decimos que *no hay correlación*. En la figura 6.3.6, a medida que *x* aumenta, *y* cambia en forma curva. Esto se explicará más tarde.

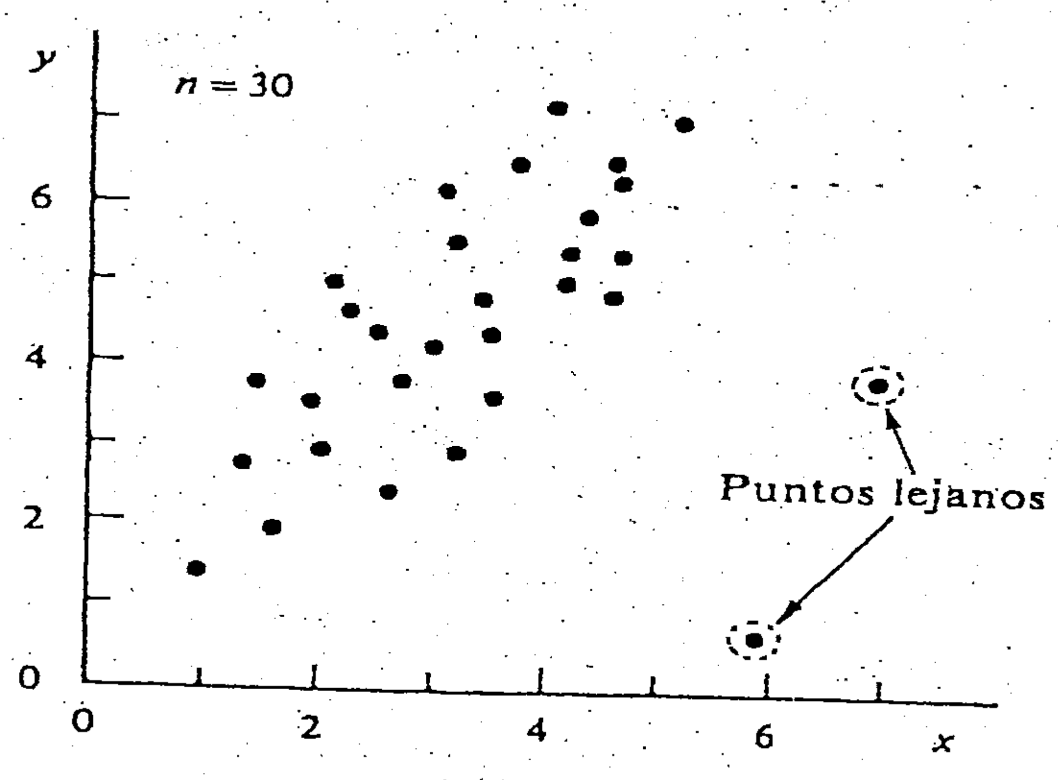


Figura 6.2 Puntos lejanos

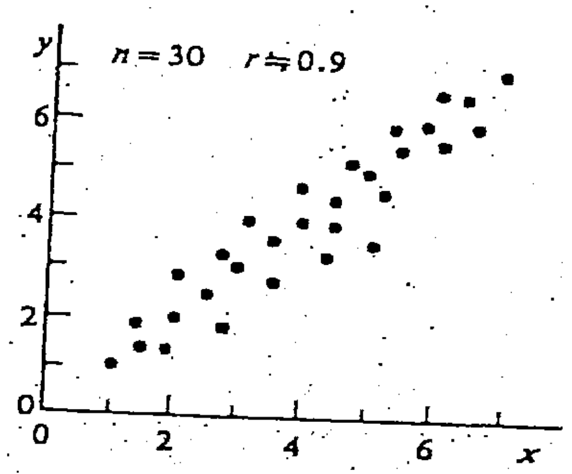


Figura 6.3.1 Correlación positiva

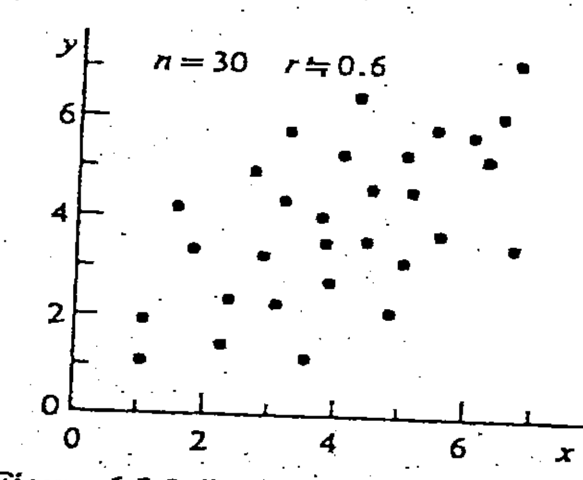


Figura 6.3.2 Puede haber correlación positiva

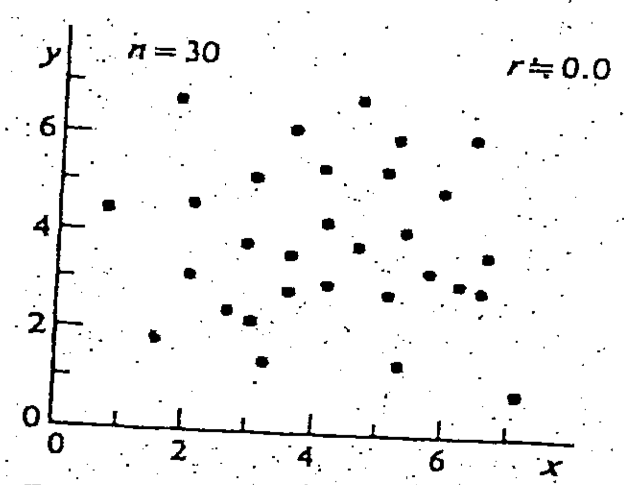


Figura 6.3.3 No hay correlación

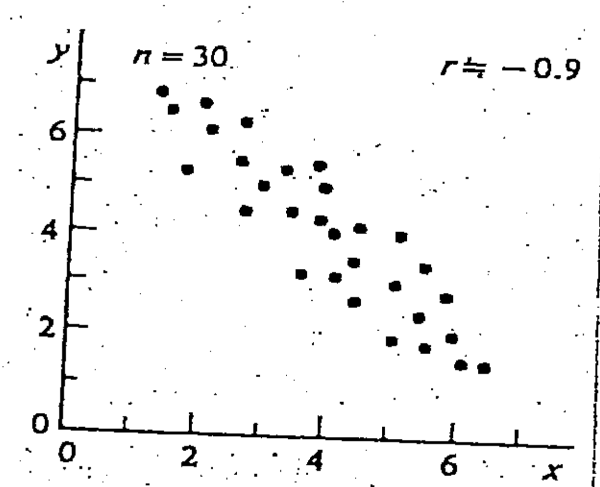


Figura 6.3.4 Correlación negativa

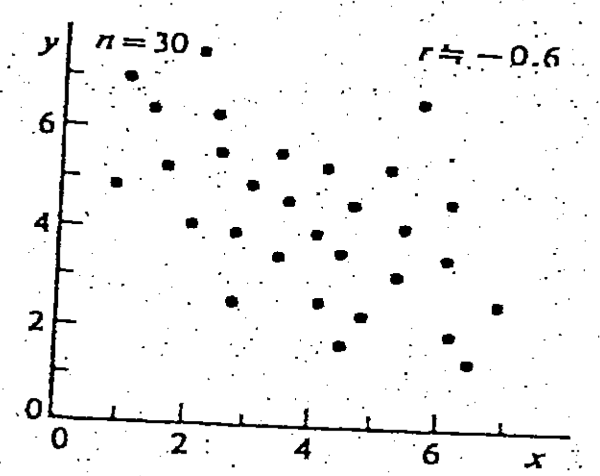


Figura 6.3.5 Puede haber correlación negativa

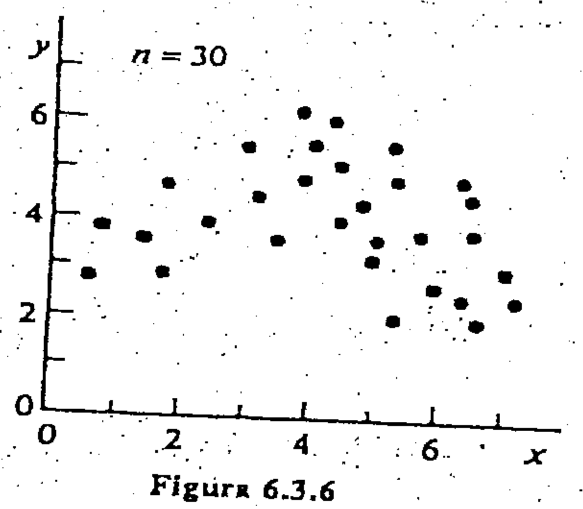


Figura 6.3.6

6.4 EL CÁLCULO DE LOS COEFICIENTES DE CORRELACIÓN

Para estudiar la relación entre x y y es importante hacer primero un diagrama de dispersión; sin embargo, para comprender la fuerza de la relación en términos cuantitativos, es útil calcular el *coeficiente de correlación* según la siguiente definición:

$$r = \frac{S(xy)}{\sqrt{S(xx) \bullet S(yy)}} \tag{6.1}$$

donde

$$S(xx) = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}, \tag{6.2}$$

$$S(yy) = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}, \tag{6.3}$$

$$S(xy) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i) \bullet (\sum_{i=1}^{n} y_i)}{n}. \tag{6.4}$$

"n" es el número de pares de datos, y $S(xy)$ se llama la *covarianza*. El coeficiente de correlación r se encuentra en el rango $-1 \leq r \leq 1$. Si el valor absoluto de r es mayor que 1, claramente ha ocurrido un error de cálculo, y debe calcularse de nuevo. En el caso de una correlación positiva fuerte como en la figura 6.3.1, se obtiene un valor cercano a +1, e igualmente, con una correlación negativa fuerte como en la figura 6.3.4, se obtiene un valor cercano a −1. Es decir, cuando $|r|$ está cerca de 1, indica una correlación fuerte entre x y y, y cuando $|r|$ está cerca de 0, una correlación débil. Además, cuando $|r| = 1$, los datos aparecerán en línea recta. Si usted recuerda esto, y adquiere el hábito



de estimar el valor de r a partir del diagrama de dispersión, podrá verificar errores de cálculo.

Calculemos el coeficiente de correlación para el ejemplo anterior de los tanques plásticos. En la tabla 6.2 se encuentra una tabla suplementaria de cálculos. A partir de esto, tenemos:

$$S(xx) = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 2312.02 - \frac{263.2^2}{30} = 2.88, \quad (6.5)$$

$$S(yy) = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 23.97833 - \frac{26.816^2}{30} \quad (6.6)$$

$$S(xy) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} \quad (6.7)$$

$$= 235.3570 - \frac{263.2 \times 26.816}{30} = 0.0913, \quad (6.8)$$

$$r = \frac{0.0913}{\sqrt{2.88 \times 0.00840}} = 0.59.$$

El valor de r es 0.59, de manera que sí hay una correlación positiva entre la presión del aire de soplado y el porcentaje de tanques plásticos defectuosos.



Fecha	x	y	x^2	y^2	xy
Oct. 1	8.6	0.889	73.96	0.79032	7.6454
2	8.9	0.884	79.21	0.78146	7.8676
3	8.8	0.874	77.44	0.76388	7.6912
4	8.8	0.891	77.44	0.79388	7.8408
5	8.4	0.874	70.56	0.76388	7.3416
8	8.7	0.886	75.69	0.78500	7.7082
9	9.2	0.911	84.64	0.82992	8.3812
10	8.6	0.912	73.96	0.83174	7.8432
11	9.2	0.895	84.64	0.80102	8.2340
12	8.7	0.896	75.69	0.80282	7.7952
15	8.4	0.894	70.56	0.79924	7.5096
16	8.2	0.864	67.24	0.74650	7.0848
17	9.2	0.922	84.64	0.85008	8.4824
18	8.7	0.909	75.69	0.82628	7.9083
19	9.4	0.905	88.36	0.81902	8.5070
22	8.7	0.892	75.69	0.79566	7.7604
23	8.5	0.877	72.25	0.76913	7.4545
24	9.2	0.885	84.64	0.78322	8.1420
25	8.5	0.866	72.25	0.74996	7.3610
26	8.3	0.896	68.89	0.80282	7.4368
29	8.7	0.896	75.69	0.80282	7.7952
30	9.3	0.928	86.49	0.86118	8.6304
31	8.9	0.886	79.21	0.78500	7.8854
Nov. 1	8.9	0.908	79.21	0.82446	8.0812
2	8.3	0.881	68.89	0.77616	7.3123
5	8.7	0.882	75.69	0.77792	7.6734
6	8.9	0.904	79.21	0.81722	8.0456
7	8.7	0.912	75.69	0.83174	7.9344
8	9.1	0.925	82.81	0.85562	8.4175
9	8.7	0.872	75.69	0.76038	7.5864
Total	263.2	26.816	2312.02	23.97833	235.3570

Tabla 6.2

6.5 NOTAS SOBRE EL ANÁLISIS CORRELACIONAL

El método de juzgar la existencia de correlación haciendo un diagrama de dispersión y calculando el coeficiente de correlación, como lo hemos descrito antes, se llama *análisis correlacional.* A continuación, en las secciones (1) a (4) se presentan algunos comentarios sobre el análisis correlacional.

(1) Coordenadas de los ejes

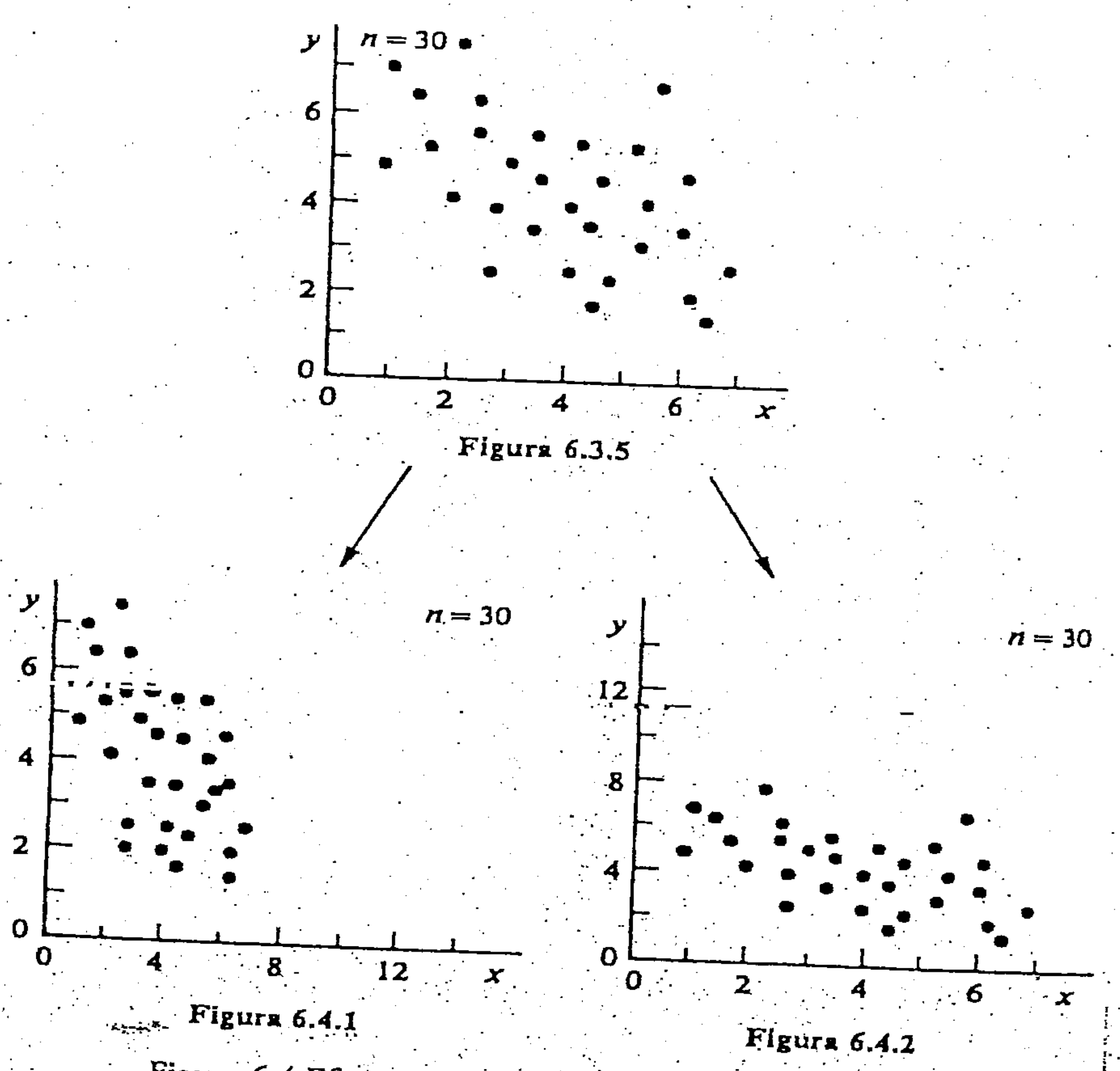


Figura 6.4 Efecto visual de las escalas de las coordenadas

En el paso 2 del dibujo del diagrama de dispersión se dieron las explicaciones respecto a la manera de dibujar las coordenadas de los ejes. En la figura 6.4 se registraron los mismos pares de datos. La escala del eje horizontal de la figura 6.4.1 es la mitad de la de la figura 6.3.5, y el eje vertical de la figura 6.4.2 es la mitad del de la figura 6.3.5. Fue posible interpretar la distribución de la figura 6.3.5 como una correlación negativa, pero esto no es claro ni en la figura 6.4.1 ni en la figura 6.4.2; incluso puede parecer que no hubiera correlación. De esta manera, si la escala está incorrectamente definida, el resultado será la interpretación errónea de los datos. Por tanto, debemos dibujar las coordenadas de los ejes en forma correcta, como se explicó en el paso 2.

(2) Estratificación

Tanto la figura 6.5.1 como la figura 6.5.2 muestran en diagramas de dispersión la relación entre la cantidad de impureza y la viscosidad de una sustancia manufacturada. En la figura 6.5.1, cuando se combinan indiscriminadamente los datos de la Compañía *A* con los de la Compañía *B*, parece no haber correlación (figura 6.5.1 a la izquierda); pero cuando se distinguen como en la figura 6.5.1 a la derecha, se ve que hay una clara correlación. Por otra parte, la figura 6.5.2 parece mostrar una correlación general, pero cuando se estratifica en *A* y *B* (figura 6.5.2 a la derecha), la corre-

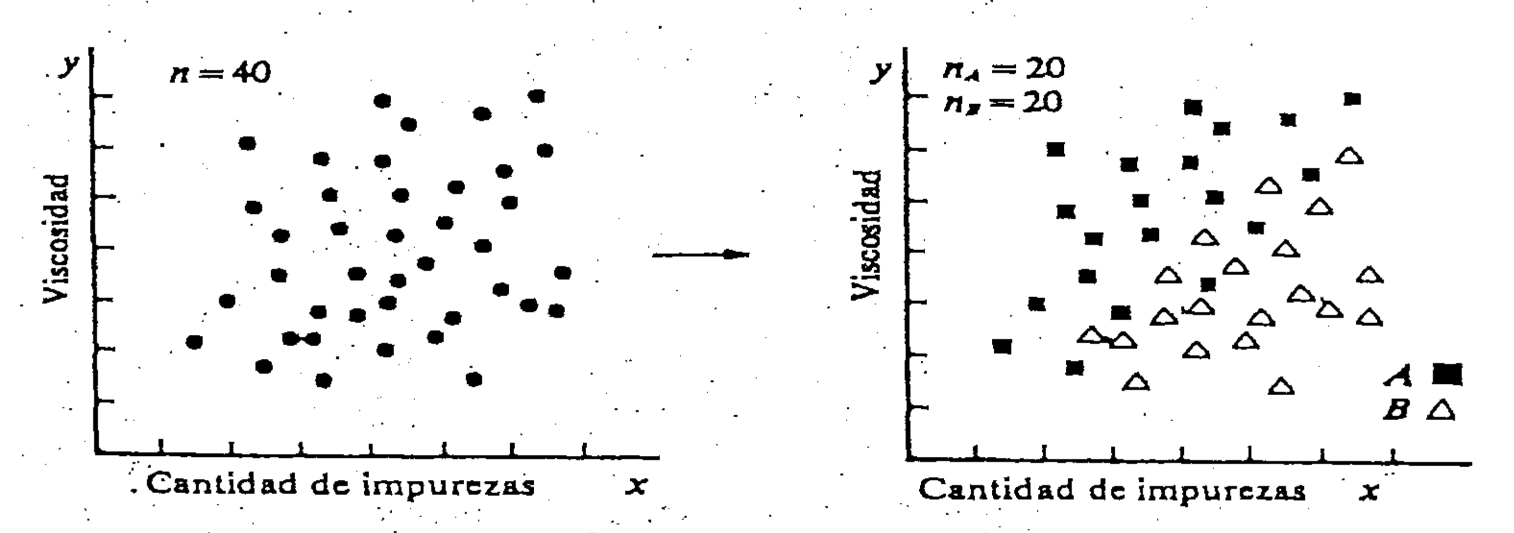


Figura 6.5.1 La estratificación en los diagramas de dispersión

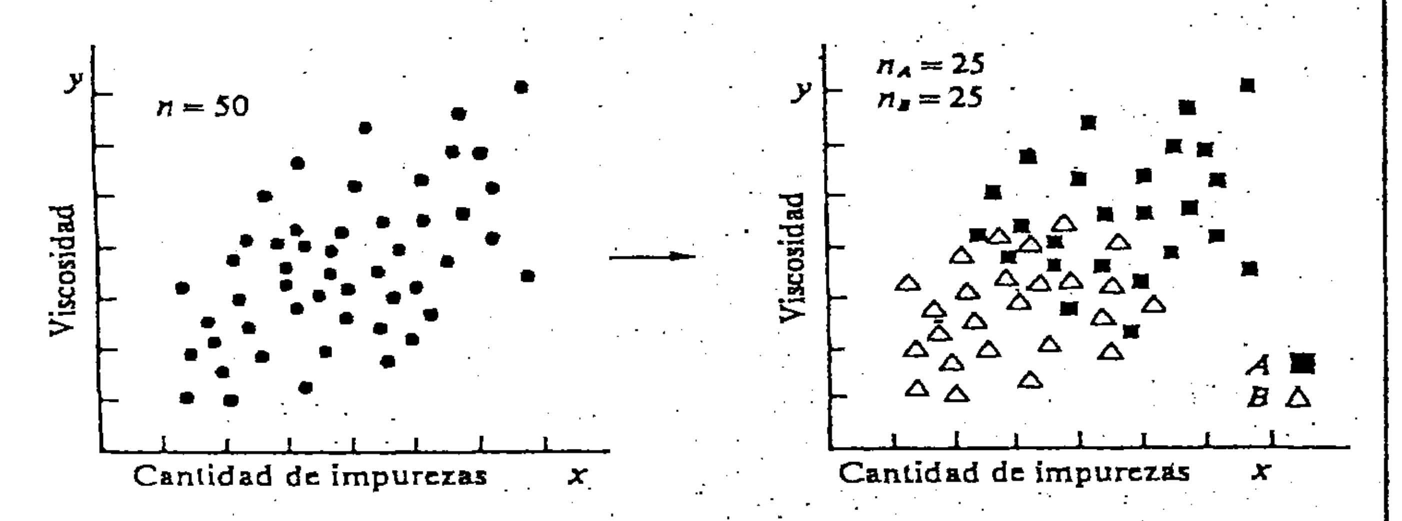


Figura 6.5.2 La estratificación en los diagramas de dispersión

lación desaparece. Cuando se ha estratificado un factor, se puede obtener información vital distinguiendo los datos de los diversos estratos por medio de colores o de símbolos. Con este fin, se debe tener cuidado siempre de registrar el origen y contexto de la información, de tal forma que los datos sean información útil en el análisis posterior.

(3) Rangos de las variables

La figura 6.6.1 y la figura 6.6.2 son dos mitades de la figura 6.3.6, dividida tomando $0 \leq x \leq 4$ para la figura 6.6.1 y $4 < x \leq 8$ para la figura 6.6.2. En la sección 6.3 se anotó que a medida que x aumenta, y cambia en forma curva. Si se examinara esto con $0 \leq x \leq 4$, probablemente se diría que hay una correlación positiva, como se ve en la figura 6.6.1. Por otra parte, en el caso de la figura 6.6.2, se vería una correlación negativa.

La figura 6.7.1 es la parte de la figura 6.3.2 desde $3 \leq x \leq 6$ solamente. En la figura 6. 3. 2 había una clara correlación positiva; pero en la figura 6.7.1 no había ninguna correlación notoria. Como se ve en estos ejemplos, analizar si hay o no una correlación depende en gran medida del rango de las variables y no se presenta necesariamente igual en todos los rangos. Lo mismo sucede en el caso de estratos diferentes. La figura 6.5.2 muestra la relación entre la cantidad de impurezas y la viscocidad con dos



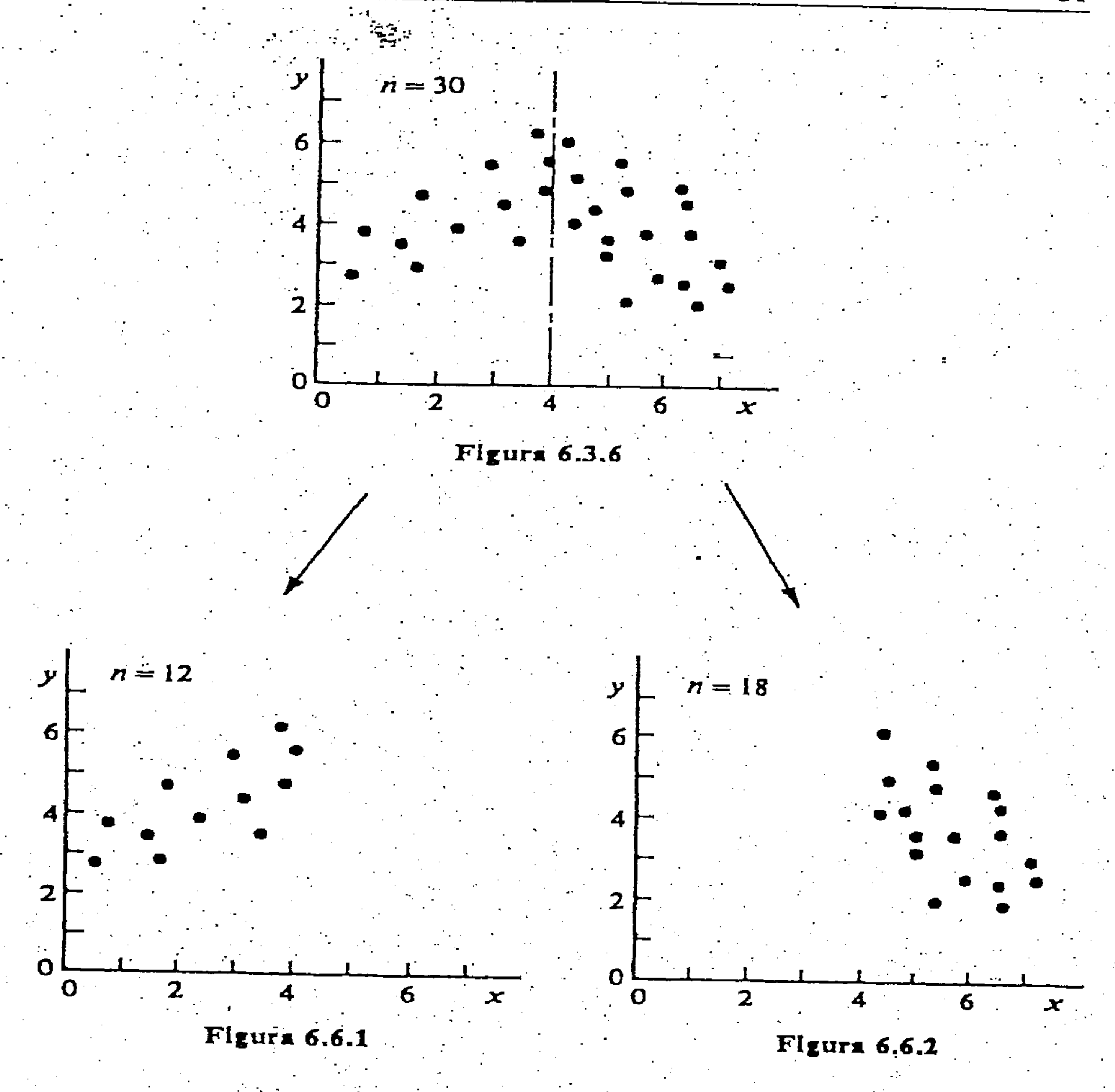


Figura 6.6 Efecto del rango de la variable (1)

lotes de *A* y *B*. Cuando el lote *A* y el lote *B* se mezclan hay una correlación, pero cuando se separan, no hay correlación. Debido a que es sumamente peligroso extrapolar una conclusión más allá de los datos obtenidos, en este caso se debe verificar por medio de experimentos o realizar una investigación técnica adecuada.

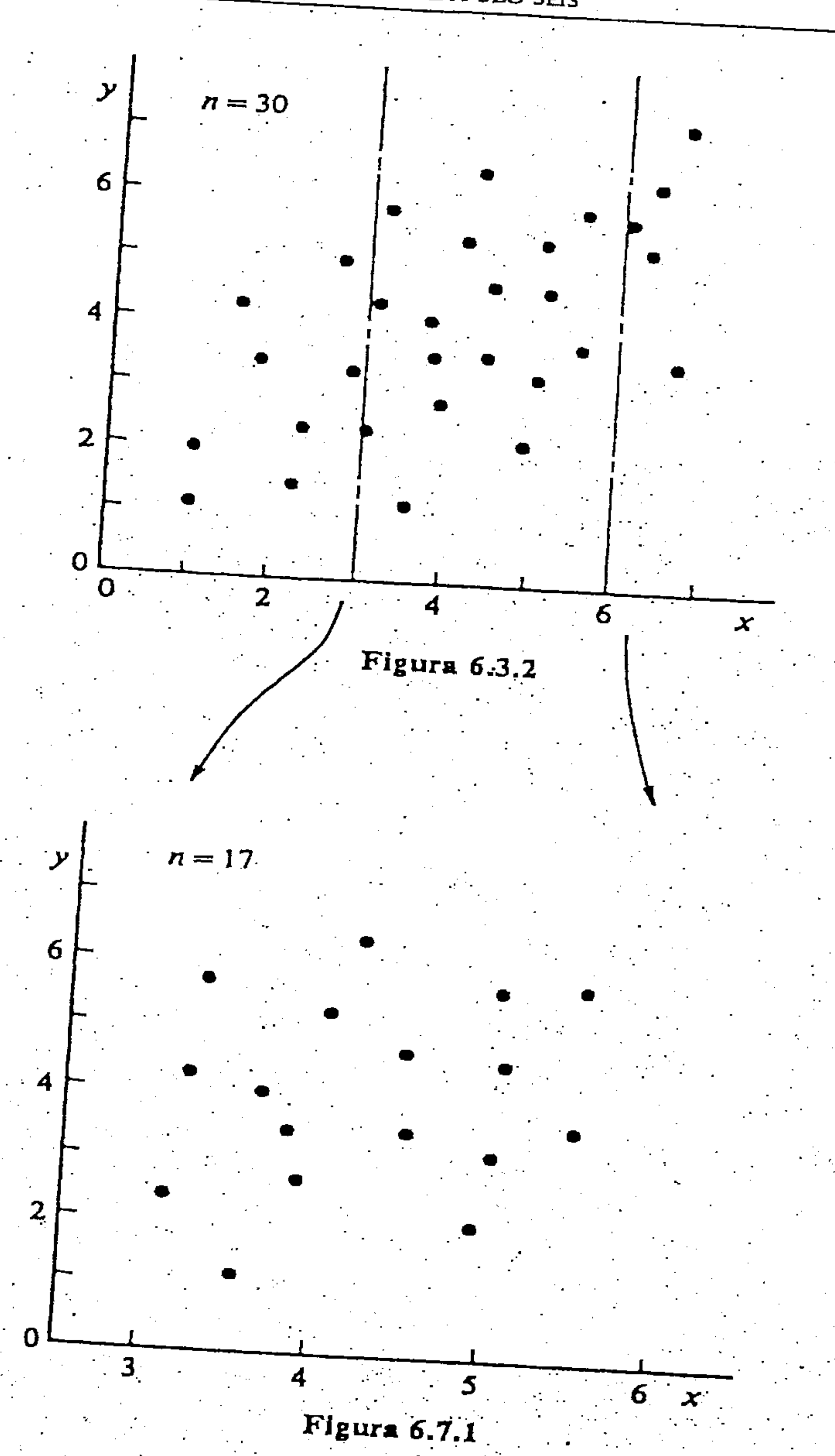


Figura 6.3.2

Figura 6.7.1

Figura 6.7 Efecto del rango de la variable (2)

(4) Correlaciones falsas:

De acuerdo con cierto estudio, existía una correlación positiva fuerte entre el índice de precios al consumidor y el número de incendios. Si esto es así, entonces ¿descenderá el número de incendios si desciende el índice de precios al consumidor? Lo más probable es que no. Para reducir el número de incendios, enfatizaríamos la importancia de limpiar los ceniceros y de no botar basura que pueda incendiarse. Así, cuando se calcula un coeficiente de correlación entre dos variables, ocurre a veces que, por casualidad, hay una alta correlación entre variables que no tienen o tienen muy poca relación de causa-efecto. Este tipo de correlación se llama una *falsa correlación*. Aun si el coeficiente de correlación es alto, no indica necesariamente una relación de causa-efecto. Es necesario tener en cuenta este hecho, y pensar en su significado científico y tecnológico.

6.6 ¿QUÉ ES EL ANÁLISIS DE REGRESIÓN?

En el ejemplo anterior de los tanques plásticos con paredes delgadas defectuosas, encontramos que existía una correlación positiva entre la presión del aire de soplado y el porcentaje de tanques defectuosos. Para evitar el problema de las paredes delgadas, debemos llevar nuestro análisis un paso más adelante. Cuando la presión del aire de soplado está en un cierto valor, ¿qué tan delgadas serán las paredes que se formarán? ¿Cómo debe controlarse la presión del aire para que las paredes no se adelgacen? Para realizar este análisis y responder las preguntas, es necesario comprender cuantitativamente la relación entre la presión del aire de moldeado y el espesor de las paredes.

La tabla 6.3 muestra los datos de un experimento en el cual la presión del aire se cambió y el espesor de las paredes de los tanques se midió en cada ocasión. La figura 6.8 es un diagrama de dispersión basado en los datos. A partir de este diagrama, parecería que la presión del aire y el espesor de las paredes tuvieran una relación en línea recta. Ahora, llamemos x a la presión del



aire y y al espesor de las paredes, y supongamos una relación en línea recta;

$$y = \alpha + \beta x$$

Este tipo de línea recta suele llamarse *línea de regresión*, donde y es la variable que responde (o variable dependiente), y x es la variable explicativa (o independiente). Así mismo, α se llama una constante y β se llama un *coeficiente de regresión*. La forma cuantitativa de captar la relación entre x y y buscando una forma de regresión de x y y se llama *análisis de regresión*.

Presión del aire (kg/cm^2)	8.0	8.5	9.0	9.5	10.0
Ancho de la pared (mm)	4.62	4.12	3.21	2.86	1.83
	4.50	3.88	3.05	2.53	2.02
	4.43	4.01	3.16	2.71	2.24
	4.81	3.67	3.30	2.62	1.95

Tabla 6.3

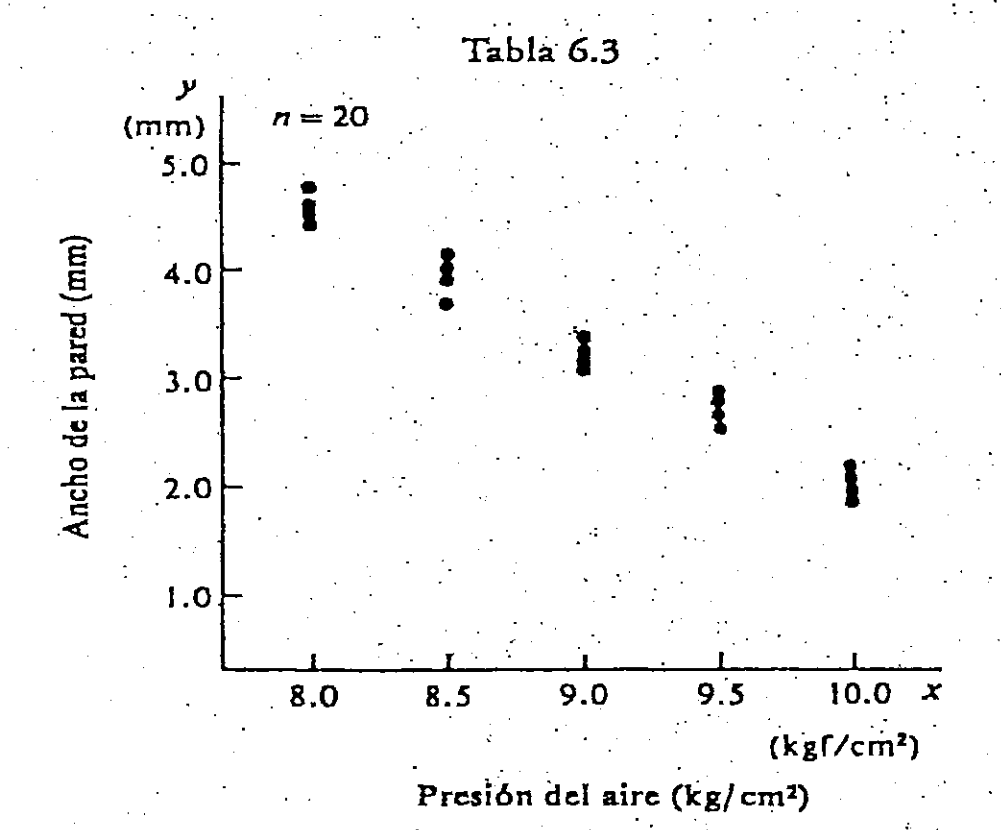


Figura 6.1 Relación entre la presióndel aire y el espesor

6.7 ESTIMACIÓN DE LAS LÍNEAS DE REGRESIÓN

Sea $(x_i y_i)$ $(1 \leq i \leq n)$ un conjunto de n pares de datos observados. Sea $\hat{\alpha}$ y $\hat{\beta}$ el valor estimado de α y β y sea e_i el residuo entre y_i, y $\hat{\alpha} + \hat{\beta} x_i$, es decir

$$e_i = y_i - (\hat{\alpha} + \hat{\beta} x_i) \quad (1 \leq i \leq n). \tag{6.9}$$

Por medio del *método de los mínimos cuadrados*, se obtienen α y β como los valores que minimizan $\sum_{i=1}^{n} e_i^2$, la suma de los residuos. Para este método se siguen los pasos siguientes:

Paso 1

Obtenga $\bar{x}$ y $\bar{y}$ de los datos.

Paso 2

Calcule $S(xx)$ y $S(yy)$.

Paso 3

Se obtiene β de

$$\hat{\beta} = \frac{S(xy)}{S(xx)}, \tag{6.10}$$

y α se obtiene de

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}. \tag{6.11}$$

Los valores de α y β obtenidos en estos pasos hacen la suma de los cuadrados de los residuos mínima.

Ahora, usando los datos en la tabla 6.3, calculemos la línea de regresión:

Paso 1

$$\bar{x} = (8.0 + 8.5 + 9.0 + 9.5 + 10.0) \times 4/20 = 9.00 \quad (6.12)$$
$$\bar{y} = (4.62 + 4.50 + \dots + 1.95)/20 = 3.276 \quad (6.13)$$

Paso 2

$$S(xx) = \sum x_1^2 - (\sum x_i)^2/n = 1630 - 180^2/20 = 10.0 \quad (6.14)$$
$$S(xy) = \sum x_i y_i - (\sum x_i) \bullet (\sum y_i)/n$$
$$= 576.88 - 180 \times 65.52/20 = -12.8 \quad (6.15)$$

Paso 3

$$\hat{\beta} = -12.8/10.0 = -1.28 \quad (6.16)$$
$$\hat{\alpha} = 3.276 - (-1.28) \times 9.00 = 14.80 \quad (6.17)$$

De esta manera, la línea de regresión se expresa por $y = 14.80 - 1.28x$. Es decir, por cada 1 kgf/cm^2 que aumente la presión del aire, el espesor de la pared disminuye en 1.28 m.m.

La figura 6.9 muestra la línea de regresión antes de calculada. La mayoría de los puntos en el diagrama de dispersión deben distribuirse alrededor de la línea de regresión. Si no, puede haber un error de cálculo y deben verificarse los pasos.

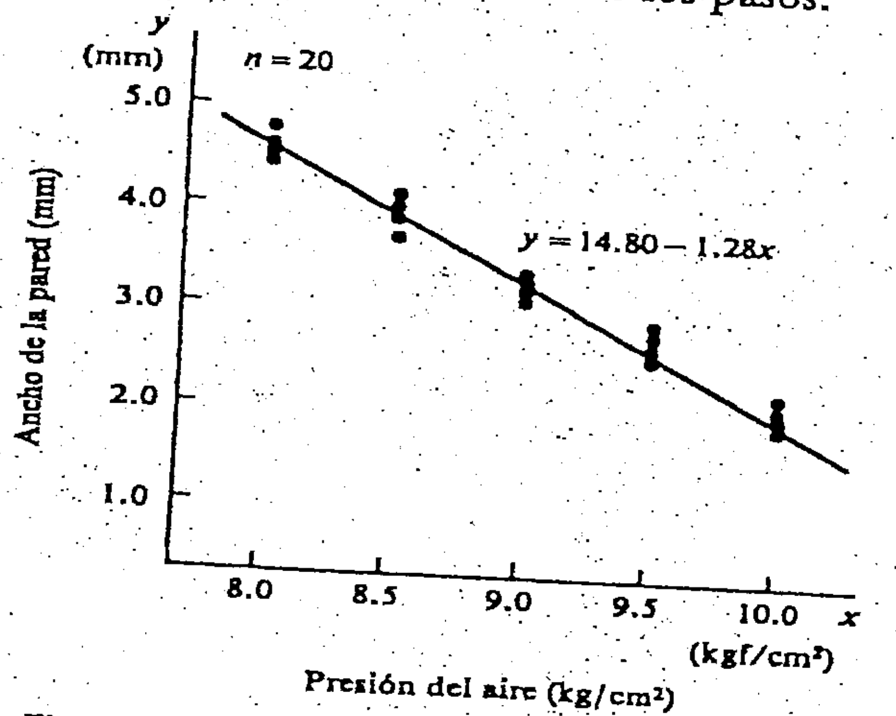


Figura 6.9 Relación entre presión del aire y espesor

6.8 NOTAS SOBRE EL ANÁLISIS DE REGRESIÓN

Algunos de los comentarios sobre diagramas de dispersión y correlaciones también pueden aplicarse al análisis de correlación. Lo que es particularmente importante es que nunca se haga un análisis de regresión sin haber hecho previamente un diagrama de dispersión.

Observe los diagramas de dispersión en las figuras 6.10.1 a 6.10.4 (figura 6.10). Ellos representan cuatro conjuntos de datos. Estos cuatro diagramas, tomados del libro de F. J. Anscombe titulado *Gráficas en el análisis estadístico*, dan casi el mismo resultado cuando se someten a análisis de regresión (ver tabla 6.4). Sin embargo, usted puede ver que la distribución de los puntos en los diagramas de dispersión correspondientes son muy diferentes entre sí. El diagrama de la figura 6.10.1 parece como si casi pudiera usarse como línea de regresión, pero en la figura 6.10.2 debe ajustarse una línea curva pues sería poco natural aplicar una línea recta. También en la figura 6.10.3 hay un punto lejano, de manera que este punto se debe dejar por fuera del análisis, o se debe repetir la medición. De los 11 conjuntos de datos en las figuras 6.10.4 el punto $x = 19$ tiene gran influencia en la determinación de la línea, pero como solamente hay información para $x = 8$ ó 19, sería mejor obtener más información. Este tipo de fenómeno sólo se puede comprender con claridad si se dibuja un diagrama de dispersión. Aunque se puede calcular una línea de regresión para cualquier tipo de datos, es imprudente calcular una línea de regresión sin dibujar un diagrama de dispersión.

El diagrama de dispersión es el punto de partida para el análisis de correlaciones y para el análisis de regresión. No olvide que el primer paso es siempre examinar cuidadosamente los datos.



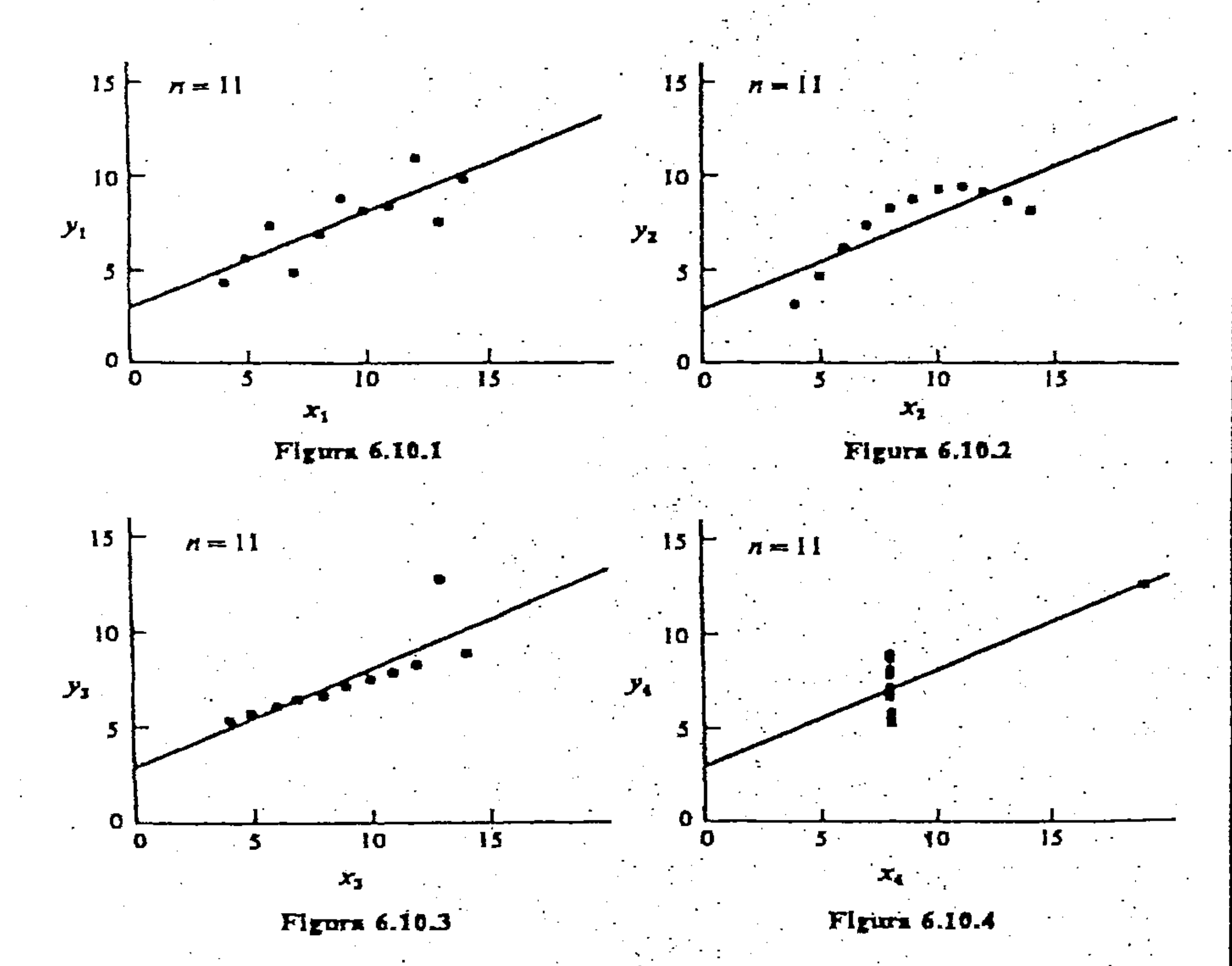


Figura 6.10 Varios diagramas de dispersión que tienen la misma línea de regresión.

No.	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.10	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.10	4	5.39	19	12.50
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89
$\bar{x}$		9.0		9.0		9.0		9.0
$\bar{y}$		7.50		7.50		7.50		7.50
$S(xx)$		110.0		110.0		110.0		110.0
$S(yy)$		41.27		41.27		41.23		41.23
$S(xy)$		55.01		55.00		54.97		54.99

• Anscombe, F. J., Graphs in Statistical Analysis, *American Statistician*, 27, 17-21 (1973)

Tabla 6.4 Ejemplo

Ejercicio 6.1

Los datos a continuación muestran el contenido de carbono x (%) y la resistencia a la tracción y (kg/mm^2) de un cierto acero.

1) Elabore un diagrama de dispersión e interprételo.
2) Obtenga el coeficiente de correlación r.
3) Obtenga una línea de regresión que estime la resistencia a la tracción y a partir del contenido de carbono x.

No.	x (%)	y (kg/mm²)
1	2.0	43
2	2.4	46
3	2.2	45
4	2.3	44
5	2.5	45
6	2.8	48
7	2.2	43
8	2.7	47
9	2.4	44
10	2.3	45
11	2.0	42
12	2.2	44
13	2.6	47
14	2.1	44
15	2.5	46
16	2.7	47
17	2.1	42
18	2.6	48
19	2.4	45
20	2.1	43
21	2.3	45
22	2.2	43
23	2.3	46
24	2.4	47
25	2.3	44
26	2.4	45
27	2.6	46
28	2.5	42
29	2.6	46
30	2.4	46

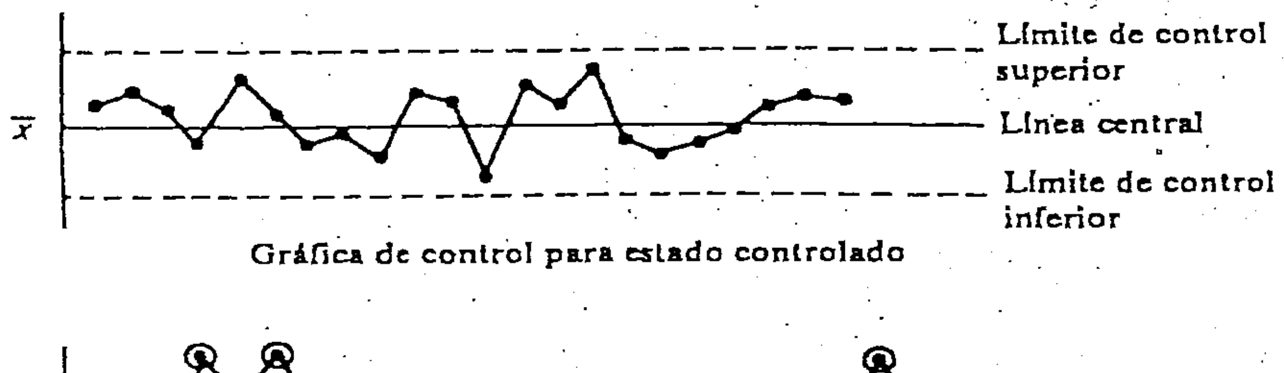
VII
Gráficas de control

7.1 ¿QUÉ SON LAS GRÁFICAS DE CONTROL

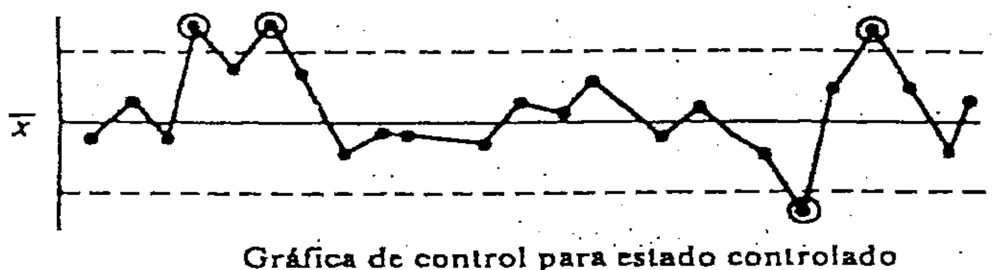
W. A. Shewhart, de los laboratorios de la Bell Telephone, fue el primero en proponer, en 1924, una gráfica de control con el fin de eliminar una variación anormal, distinguiendo las variaciones debidas a *causas asignables* de aquéllas debidas a *causas al azar*. Una gráfica de control consiste en una línea central, un par de límites de control, uno de ellos colocado por encima de la línea central, y otro por debajo, y en unos valores característicos registrados en la gráfica que representa el estado del proceso. Si todos los valores ocurren dentro de los límites de control, sin ninguna tendencia especial, se dice que el proceso está en estado controlado. Sin embargo, si ocurren por fuera de los límites de control o muestran una forma peculiar, se dice que el proceso está fuera de control. La figura 7.1 muestra algunos ejemplos.

La calidad de un producto manufacturado por medio de un proceso inevitablemente sufrirá variaciones. Estas variaciones tienen causas y éstas últimas pueden clasificarse en dos tipos: causas debidas al azar y causas asignables.





Gráfica de control para estado controlado



Gráfica de control para estado controlado

Figura 7.1 Ejemplos de gráficas de control

Causas debidas al azar

Las variaciones debidas al azar son inevitables en el proceso, aun si la operación se realiza usando materia prima y métodos estandarizados. No es práctico eliminar el azar técnicamente y en forma económica por el momento.

Causas asignables

La variación debida a causas asignables significa que hay factores significativos que pueden ser investigados. Es evitable y no se puede pasar por alto: hay casos causados por la no aplicación de ciertos estándares o por la aplicación de estándares inapropiados.

Cuando los puntos se ubican por fuera de los límites de control o muestran una tendencia particular, decimos que el proceso está *fuera de control*, y esto equivale a decir, "Existe variación por causas asignables y el proceso está en un estado de descontrol". Para controlar un proceso, se requiere poder predecir el resultado dentro de un margen de variación debido al azar.

Para hacer una gráfica de control es necesario estimar la variación debida al azar. Para esto se dividen los datos en *subgrupos* dentro de los cuales el lote de materia prima, las máquinas, los



operadores y otros factores son comunes, de modo que la variación dentro del subgrupo puede considerarse aproximadamente la misma que la variación por causas debidas al azar.

Hay varias clases de gráficas de control, dependiendo de su propósito y de las características de la variable. En cualquier tipo de gráfica de control el límite de control se calcula usando la siguiente fórmula:

$$(\text{valor promedio}) \pm 3 \times (\text{desviación estándar}),$$

donde la desviación estándar es la variación debida al azar. Este tipo de gráfica de control se llama una gráfica de control de 3-sigma.

7.2 TIPOS DE GRÁFICAS DE CONTROL

Hay dos tipos de gráficas de control, una para valores continuos y otra para valores discretos. Los tipos de gráfica prescritos por JIS* se muestran en la tabla 7.1 y sus líneas de control se muestran en la tabla 7.2

Valor característico	Nombre
Valor continuo	Gráfica $\bar{x} - R$ (Valor promedio y rango) Gráfica x (Variable de medida)
Valor discreto	Gráfica pn (Número de unidades defectuosas) Gráfica p (Fracción de unidades defectuosas) Gráfica c (Número de defectos) Gráfica u (Número de defectos por unidad)

Tabla 7.1 Tipos de gráfica de control

* JIS son las siglas de las Normas Industriales Japonesas o Japanese Industrial Standards.

Tipo de gráfica de control	Límite superior de control (LCs), Línea central (LC), Límite inferior de control (LCi)
Valor continuo – promedio $\bar{x}$	$LCs = \bar{\bar{x}} + A_2\bar{R}$ $LC = \bar{\bar{x}}$ $LCi = \bar{\bar{x}} - A_2\bar{R}$
Valor continuo – rango R	$LCs = D_4\bar{R}$ $LC = \bar{R}$ $LCi = D_3\bar{R}$
Valor continuo – valor medido x	$LCs = \bar{x} + 2.66\bar{R}_s$ $LC = \bar{x}$ $LCi = \bar{x} - 2.66\bar{R}_s$
Valor discreto – número de unidades defectuosas pn	$LCs = \bar{p}n + 3\sqrt{\bar{p}n(1-\bar{p})}$ $LC = \bar{p}n$ $LCi = \bar{p}n - 3\sqrt{\bar{p}n(1-\bar{p})}$
Valor discreto – fracción de unidades defectuosas p	$LCs = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n}$ $LC = \bar{p}$ $LCi = \bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n}$
Valor discreto – número de defectos c	$LCs = \bar{c} + 3\sqrt{\bar{c}}$ $LC = \bar{c}$ $LCi = \bar{c} - 3\sqrt{\bar{c}}$
Valor discreto – número de defectos por unidad u	$LCs = \bar{u} + 3\sqrt{\bar{u}/n}$ $LC = \bar{u}$ $LCi = \bar{u} - 3\sqrt{\bar{u}/n}$

Tabla 7.2 Lista de fórmulas para líneas de control

(1) Gráfica $\overline{x} - R$

Ésta se usa para controlar y analizar un proceso en el cual la característica de calidad del producto que se está midiendo toma valores continuos, tales como longitud, peso o concentración, y esto proporciona la mayor cantidad de información sobre el proceso. $\overline{x}$ representa un valor promedio de un subgrupo y R representa el rango del subgrupo. Una gráfica R se usa generalmente en combinación con una gráfica $\overline{x}$ para controlar la variación dentro de un subgrupo.

(2) Gráfica x

Cuando los datos de un proceso se registran durante intervalos largos o los subgrupos de datos no son efectivos, se grafica cada dato individualmente y esa gráfica puede usarse como gráfica de control. Debido a que no hay subgrupo y el valor R no puede calcularse, se usa el rango móvil Rs de datos sucesivos para el cálculo de los límites de control de x.

(3) Gráfica pn, Gráfica p

Estas gráficas se usan cuando la característica de calidad se representa por el número de unidades defectuosas o la fracción defectuosa. Para una muestra de tamaño constante, se usa una gráfica pn del número de unidades defectuosas, mientras que una gráfica p de la fracción de defectos se usa para una muestra de tamaño variable.

(4) Gráfica c, Gráfica u

Éstas se usan para controlar y analizar un proceso por los defectos de un producto, tales como rayones en placas de metal, número de soldaduras defectuosas de un televisor o tejido desigual en telas. Una gráfica c referida al número de defectos se usa para un producto cuyas dimensiones son constantes, mientras que una gráfica u se usa para un producto de dimensión variable.



7.3 CÓMO ELABORAR UNA GRÁFICA DE CONTROL

(1) Gráfica $\bar{x} - R$

Procedimiento	Ejemplo
Paso 1 Recoja los datos Recoja aproximadamente 100 datos. Divídalos en 20 ó 25 subgrupos con 4 ó 5 en cada uno, haciéndolos uniformes dentro del subgrupo. Regístrelos en una hoja de datos (tabla 7.3). Cuando no hay razones técnicas para hacer subgrupos, divida los datos en el orden en que se obtuvieron. El tamaño del grupo es generalmente entre 2 y 10 en la mayoría de los casos.	
Paso 2 Calcule los $\bar{x}$'s Calcule el promedio $\bar{x}$ para cada subgrupo. $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$ donde n es el tamaño de cada subgrupo. Por lo general, el resultado se calcula con una cifra decimal más que aquéllas de los datos originales.	*Paso 2* En el primer grupo, $\bar{x} = (47+32+44+35+20)/5 = 35.6$



Paso 3 Calcule $\overline{\overline{x}}$ *Paso 3*

Calcule el promedio bruto $\overline{\overline{x}}$ dividiendo el total de los $\overline{x}$ s de cada subgrupo por el número de subgrupos k.

$$\overline{\overline{x}} = \frac{\overline{x}_1 + \overline{x}_2 + \cdots + \overline{x}_n}{k}$$

$\overline{\overline{x}} = (35.6+29.2+\cdots+28.2)/25 = 29.86$

$\overline{\overline{x}}$ se calcula con dos cifras decimales más que aquéllas de los datos originales.

Subgrupo No.	x_1	x_2	x_3	x_4	x_5	Σx	$\overline{X}$	R
1	47	32	44	35	20	178	35.6	27
2	19	37	31	25	34	146	29.2	18
3	19	11	16	11	44	101	20.2	33
4	29	29	42	59	38	197	39.4	30
5	28	12	45	36	25	146	29.2	33
6	40	35	11	38	33	157	31.4	29
7	15	30	12	33	26	116	23.2	21
8	35	44	32	11	38	160	32.0	33
9	27	37	26	20	35	145	29.0	17
10	23	45	26	37	32	163	32.6	22
11	28	44	40	31	18	161	32.2	26
12	31	25	24	32	22	134	26.8	10
13	22	37	19	47	14	139	27.8	33
14	37	32	12	38	30	149	29.8	26
15	25	40	24	50	19	158	31.6	31
16	7	31	23	18	32	111	22.2	25
17	38	0	41	40	37	156	31.2	41
18	35	12	29	48	20	144	28.8	36
19	31	20	35	24	47	157	31.4	27
20	12	27	38	40	31	148	29.6	28
21	52	42	52	24	25	195	39.0	28
22	20	31	15	3	28	97	19.4	28
23	29	47	41	32	22	171	34.2	25
24	28	27	22	32	54	163	32.6	32
25	42	34	15	29	21	141	28.2	27
Total							746.6	686
Promedio							$\overline{\overline{x}} =$ 29.86	$\overline{R} =$ 27.44

Tabla 7.3 Hoja de datos para una gráfica $\overline{x} - R$

Paso 4 Calcule R

Calcule el rango de cada subgrupo R, restando el valor mínimo del valor máximo de los datos en el subgrupo.
R = (valor máximo en un subgrupo) – (valor mínimo en un subgrupo)

Paso 4

Para el primer grupo,

$R = 47 - 20 = 27$

Paso 5 Calcule $\overline{R}$

Calcule el promedio $\overline{R}$ del rango R, dividiendo el total de los Rs de cada subgrupo por el número de grupos k.

$$\overline{R} = \frac{R_1 + R_2 + \cdots + R_k}{k}$$

$\overline{R}$ debe calcularse con dos cifras decimales más que aquéllas de los datos originales (el mismo número de decimales que el de $\overline{\overline{x}}$).

Paso 5

$\overline{R} = (27+18+\cdots+27)/25$
$= 27.44$

Paso 6 Calcule las líneas de control

Calcule cada una de las líneas de control para la gráfica $\overline{x}$ y la gráfica R con las siguientes fórmulas:

Paso 6

Gráfica $\bar{x}$

Línea central:

$$LC = \bar{\bar{x}}$$

Límite de control superior:

$$LCs = \bar{\bar{x}} + A_2 \bar{R}$$

Límite de control inferior:

$$LCi = \bar{\bar{x}} - A_2 \bar{R}$$

Gráfica $\bar{x}$

$$LC = \bar{\bar{x}} = 29.86$$

$$LCs = \bar{\bar{x}} + A_2 \bar{R} = 29.86 + 0.577 \times 27.44 = 45.69$$

$$LCi = \bar{\bar{x}} - A_2 \bar{R} = 29.86 - 0.577 \times 27.44 = 14.03$$

Gráfica R

Línea central:

$$LC = \bar{R}$$

Límite de control superior:

$$LCs = D_4 \bar{R}$$

Límite de control inferior:

$$LCi = D_3 \bar{R}$$

LCi no se tiene en cuenta cuando n es menor que 6.

A_2, D_4 y D_3 son los coeficientes determinados por el tamaño de un subgrupo (n), y se muestran en la tabla 7.4 y también en la tabla A.2 de los anexos.

Gráfica R

$$LC = \bar{R} = 27.44$$

$$LCs = D_4 \bar{R} = 2.115 \times 27.44 = 58.04$$

LCi = – (no se considera)

Tamaño del subgrupo n	Gráfica $\overline{x}$	Gráfica R		
	A_2	D_3	D_4	d_2
2	1.880	—	3.267	1.128
3	1.023	—	2.575	1.693
4	0.729	—	2.282	2.059
5	0.577	—	2.115	2.326
6	0.483	—	2.004	2.534

Tabla 7.4 Lista de coeficientes para gráficas $\overline{x} - R$

Paso 7 Dibuje las líneas de control

Paso 7

Primero, prepare una hoja de papel cuadriculado y marque el eje vertical de la izquierda con los valores de $\overline{x}$ y de R y el eje horizontal con el número de subgrupos. Para el eje vertical escoja una escala tal que los límites de control superior e inferior queden a una distancia de 20-30 mm uno del otro. Dibuje una línea sólida para la línea central y una línea punteada para los límites.

Paso 8 Localice los puntos

Paso 8

Registre los valores de $\overline{x}$ y de R de cada subgrupo sobre la misma línea vertical en el or-

den del número del subgrupo. Marque el número del subgrupo sobre la línea horizontal a intervalos de 2 a 5 mm. Use • para marcar las $\bar{x}$ y × para R con el fin de que se puedan reconocer fácilmente y enciérrelos en un círculo para los valores que estén por fuera de los límites.

Paso 9 Registre los datos que puedan ser de utilidad	*Paso 9*

Escriba el tamaño del subgrupo (n) en el extremo superior izquierdo de la gráfica $\bar{x}$. Incluya también cualquier otro aspecto relevante para el proceso, tal como los nombres del proceso y del producto, el período, el método de medición, las condiciones de trabajo, el turno, etc.

(2) Gráfica pn

Procedimiento	Ejemplo
Paso 1 Reúna los datos	*Paso 1*

Tome una muestra y clasifique la calidad del producto en unidades que llenen o no los requisitos, según el estándar de

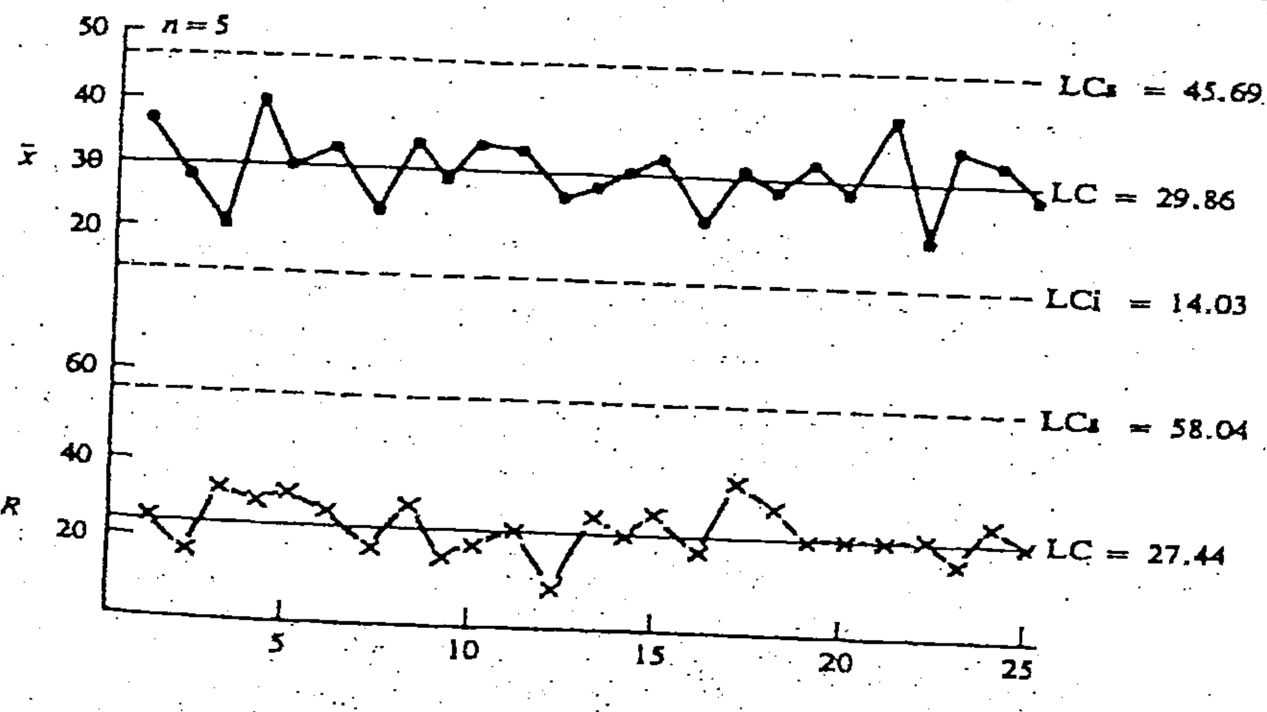


Figura 7.2 Gráfico $\bar{x} - R$

inspección. En este caso, tome una muestra de tamaño tal que la mayoría de los subgrupos tengan entre 1 y 5 unidades defectuosas, y recoja de 20 a 25 subgrupos (tabla 7.5).

Paso 2 *Calcule $\overline{p}$*

Paso 2

Calcule la fracción promedio de defectos $\overline{p}$ dividiendo el número total de unidades defectuosas de cada subgrupo por el número total de muestras.

Subgrupo No.	Tamaño del subgrupo n	(Número de unidades defectuosas) pn
1	100	4
2	100	2
3	100	0
4	100	5
5	100	3
6	100	2
7	100	4
8	100	3
9	100	2
10	100	6
11	100	1
12	100	4
13	100	1
14	100	0
15	100	2
16	100	3
17	100	1
18	100	6
19	100	1
20	100	3
21	100	3
22	100	2
23	100	0
24	100	7
25	100	3
Total	$\Sigma n = 2500$	$\Sigma pn = 68$

Tabla 7.5 Hoja de datos para gráfica pn

$\bar{p} = \frac{\sum pn}{k \times n}$	$\bar{p} = \frac{\sum pn}{k \times n} = \frac{68}{25 \times 100} = 0.0272$

Paso 3 Calcule las líneas de control	*Paso 3*
Línea central LC = $\bar{p}n$	LC = $\bar{p}n = 0.0272 \times 100 = 2.72$
Límite de control superior: LCs = $\bar{p}n + 3\sqrt{\bar{p}n(1-\bar{p})}$	LCs = $\bar{p}n + 3\sqrt{\bar{p}n(1-\bar{p})}$ $2.72 + 3\sqrt{2.72 \times (1 - 0.0272)} = 7.60$
Límite de control inferior: LCi = $\bar{p}n - 3\sqrt{\bar{p}n(1-\bar{p})}$	LCi = $\bar{p}n - 3\sqrt{\bar{p}n(1-\bar{p})}$ = $2.72 - 3\sqrt{2.72 \times (1 - 0.0272)}$
LCi no se tiene en cuenta cuando su valor es un número negativo.	No se considera

Paso 4 Construya la gráfica de control	*Paso 4*
Marque el eje horizontal con el número de subgrupos y el eje vertical con el número de unidades defectuosas. Dibuje una línea sólida para la línea central $\bar{p}n$ y líneas discontinuas para LCs y LCi. Luego, registre el número de unidades defectuosas de cada subgrupo.	

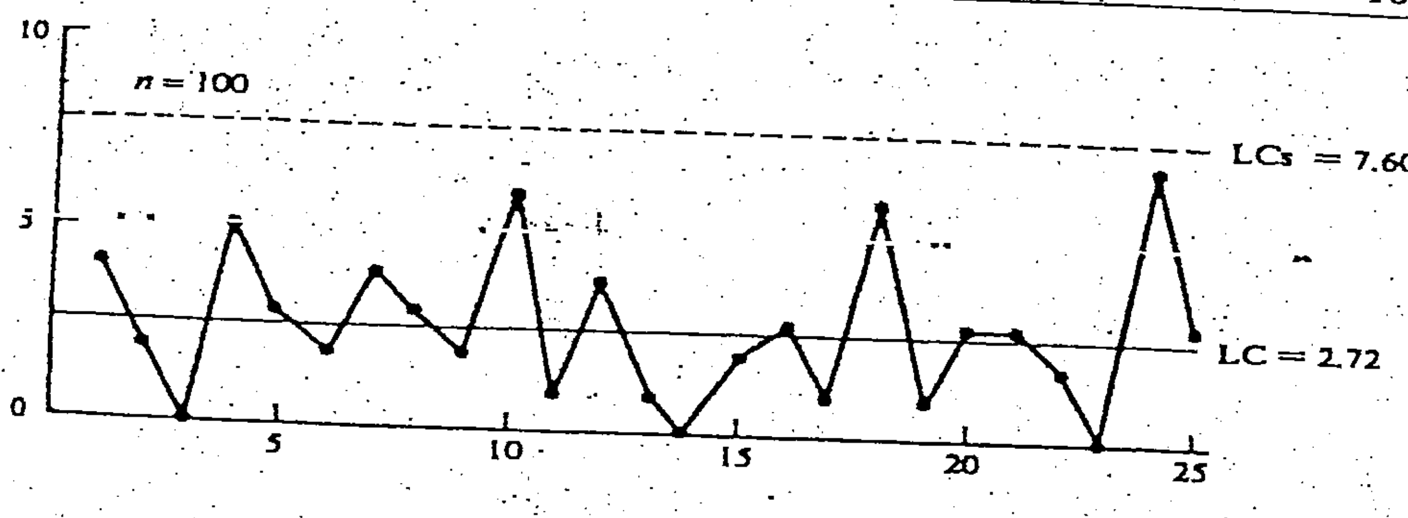


Figura 7.3 Gráfica *pn*

7.4 CÓMO LEER LAS GRÁFICAS DE CONTROL

Lo más importante en el control del proceso es captar el estado del proceso de manera precisa leyendo la gráfica de control y diligentemente tomar acciones apropiadas cuando se encuentre algo anormal en el proceso. El estado controlado del proceso es el estado en el cual el proceso es estable, es decir, el promedio y la variación del proceso no cambian. Si un proceso está o no controlado se juzga según los siguientes criterios a partir de la gráfica de control.

(1) Fuera de los límites de control

Puntos que están por fuera de los límites de control.

(2) Racha

La racha es el estado en el cual los puntos ocurren continuamente en un lado de la línea central y el número de puntos se llama longitud de la racha.

Una longitud de siete puntos en una racha se considera normal.

Aun si la longitud de la racha está por debajo de 6, se consideran anormales los siguientes casos:

a) Al menos 10 de 11 puntos consecutivos ocurren en un mismo lado de la línea central.
b) Al menos 12 de 14 puntos consecutivos ocurren en un mismo lado de la línea central.
c) Al menos 16 de 20 puntos consecutivos ocurren en un mismo lado de la línea central.

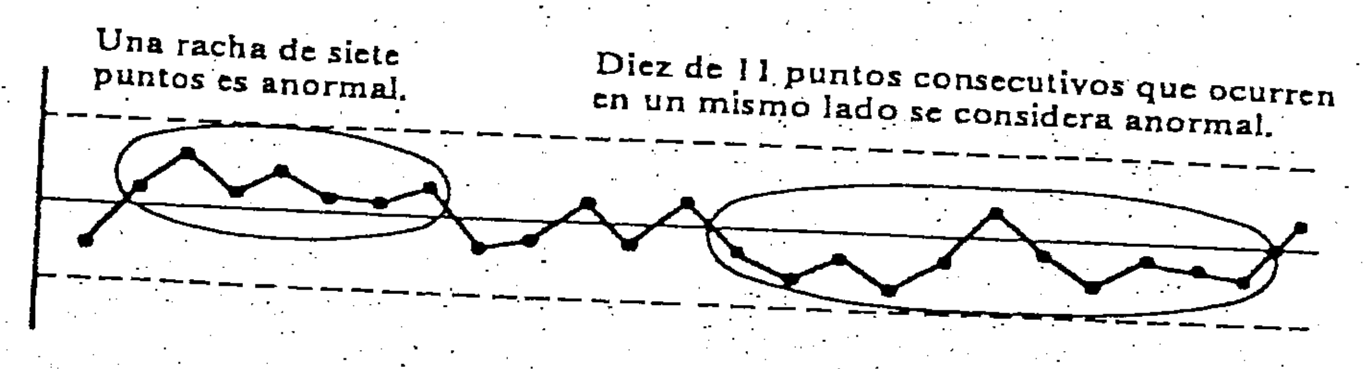


Figura 7.4.1 Racha

(3) Tendencia

Cuando los puntos forman una curva continua ascendente o descendente, se dice que hay una tendencia.

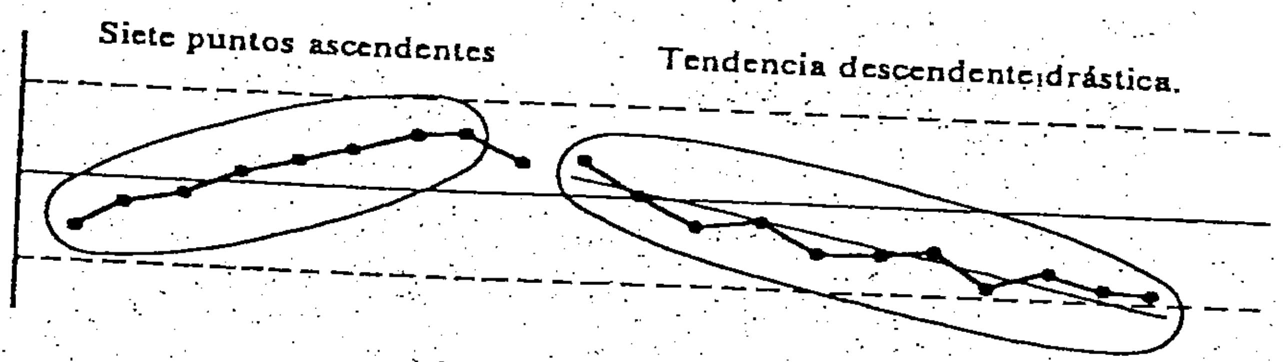


Figura 7.4.2 Tendencia

(4) Acercamiento a los límites de control

Teniendo en cuenta los puntos que se acercan a los límites de control de 3 -sigma, si 2 de 3 puntos ocurren por fuera de las líneas de 2 -sigma, el caso se considera anormal.

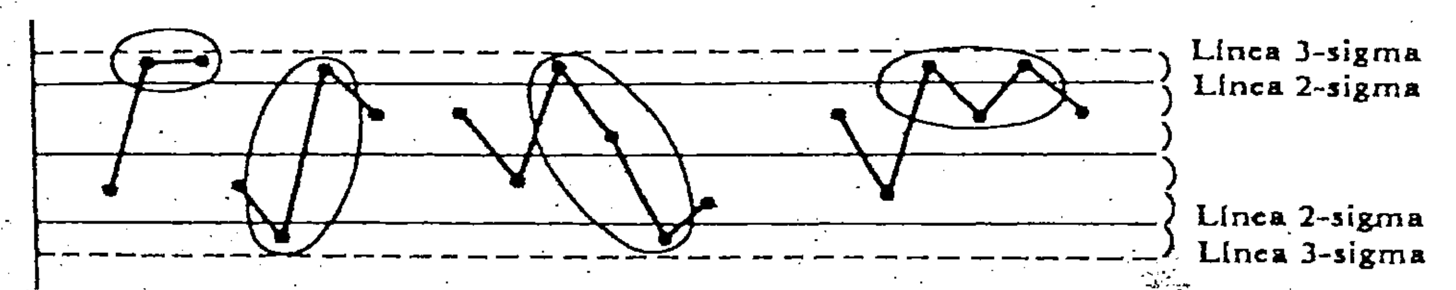


Figura 7.4.3 Acercamiento a los límites de control (2 de 3 puntos)

(5) Acercamiento a la línea central

Cuando la mayoría de los puntos están dentro de las líneas de 1.5-sigma (los bisectores de la línea central y de cada uno de los límites de control), esto se debe a una forma inapropiada de hacer los subgrupos. El acercamiento a la línea central no significa un estado de control, sino una mezcla de la información de diferentes poblaciones en los subgrupos, lo cual hace que los límites de control sean demasiado amplios. Cuando se presenta esta situación es necesario cambiar la manera de hacer los subgrupos.

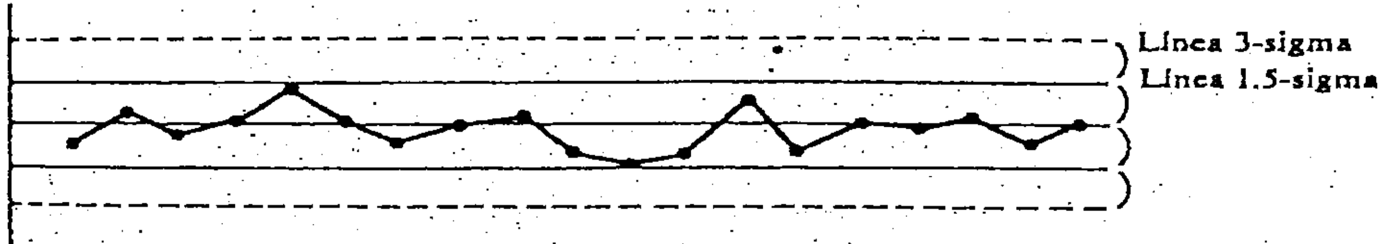


Figura 7.4.4 Acercamiento a la línea central

(6) Periodicidad

También es anormal que la curva muestre repetidamente una tendencia ascendente y descendente para casi el mismo intervalo.

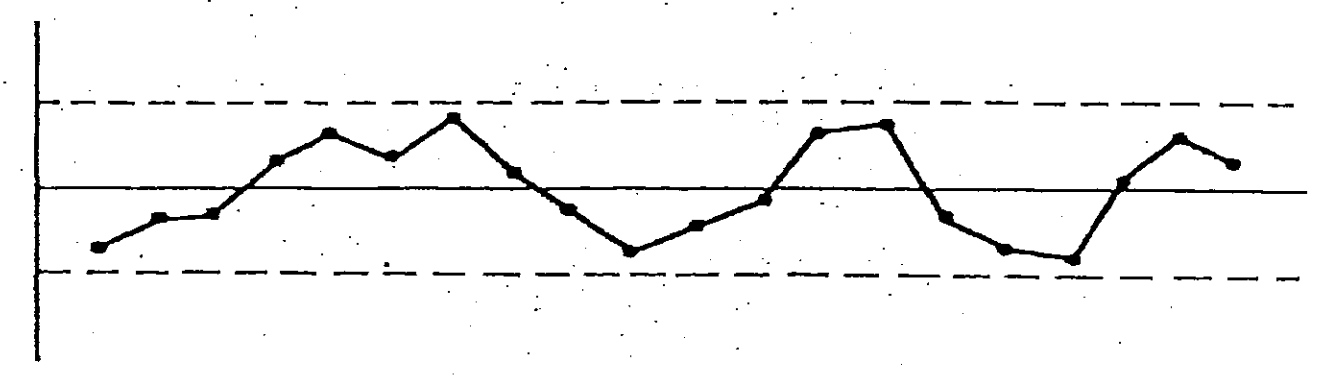


Figura 7.4.5 Periodicidad

7.5 ANÁLISIS DEL PROCESO USANDO LAS GRÁFICAS DE CONTROL

El objetivo del *análisis del proceso* puede definirse como la identificación de causas específicas asignables de la variación de una característica de calidad en un proceso. Después de encontrar esas causas asignables por medio del análisis del proceso, es necesario realizar una serie de acciones correctivas en relación con las causas asignables.

(1) Organización de los subgrupos

La organización de los subgrupos es la parte más importante de la preparación de una gráfica de control y determina su efectividad. La organización inapropiada de los subgrupos tiene como resultado una gráfica inútil.

Los datos se recogerán una vez que se haya decidido la característica de calidad de un proceso que se va a analizar. La variación en una característica de calidad en un proceso se debe a varias causas. Por esto, antes de organizar los subgrupos, es necesario considerar la variación que se ha de eliminar, y tratar luego de agrupar los datos de tal manera que la variación debida a factores permisibles constituya la variación dentro del subgrupo. Para este fin:

a) La operación debe realizarse casi en las mismas condiciones (desde el punto de vista técnico).
b) Los datos recogidos a lo largo de un período de tiempo relativamente corto deben agruparse.

En la formación de los subgrupos deben considerarse los siguientes puntos:

a) Hay maneras diferentes de agrupar. Se debe cambiar el tamaño del subgrupo y ensayar diversas maneras de combinar los datos.



b) Un cambio en la manera de organizar los subgrupos producirá un cambio en los factores que constituyen la variación dentro del subgrupo.

No puede usarse efectivamente una gráfica de control sin conocer los componentes de la variación dentro del subgrupo. El siguiente ejemplo muestra cómo los componentes de la variación dentro del subgrupo cambian, según la manera de organizar los subgrupos.

Ejemplo 7.1

En un proceso mecánico, se usa una dimensión de una parte como control de la característica del proceso. Los factores más importantes que afectan a la dimensión son la calidad de las materias primas, la forma de la herramienta de corte y el ajuste de las herramientas para la abrasión. Las condiciones del proceso son las siguientes:

a) Un lote de materia prima corresponde al trabajo de una semana.
b) Se afilan las herramientas de corte todos los días.
c) Las herramientas deben ajustarse al comenzar el turno de la mañana y el de la tarde.

En este caso se considera admisible la variación dentro de un ajuste y el criterio para el control será si el ajuste se hace correctamente y si las herramientas están en condiciones normales. Entonces los datos deben agruparse de tal manera que la variación dentro de un ajuste se convierta en la variación dentro del subgrupo y la variación entre ajustes se convierta en la variación entre los subgrupos.

La tabla 7.6 muestra la hoja de datos, las tablas 7.7 y 7.8 muestran la manera de formar los subgrupos, y la tabla 7.9 muestra los componentes de las variaciones dentro del subgrupo y entre subgrupos, respectivamente.

Fecha		Datos *n*		Comentarios		
Dic. 1	a.m.	39	39	Ajuste	Herramienta No. 1	Lote de material No. 1
	p.m.	40	42	Ajuste		
Dic. 2	a.m.	42	41	Ajuste	Herramienta No. 2	
	p.m.	43	42	Ajuste		
Dic. 3	a.m.	38	37	Ajuste	Herramienta No. 3	
	p.m.	38	38	Ajuste		

Tabla 7.6 Hoja de datos

Fecha		Datos	
Dic. 1	a.m.	39	39
	p.m.	40	42
Dic. 2	a.m.	42	41
	p.m.	43	42
Dic. 3	a.m.	38	37
	p.m.	38	38

Tabla 7.7 Subgrupos de $n = 2$

Fecha		Datos	
Dic. 1	a.m.	39	39
	p.m.	40	42
Dic. 2	a.m.	42	41
	p.m.	43	42
Dic. 3	a.m.	38	37
	p.m.	38	38

Tabla 7.8 Subgrupos de $n = 4$

Variación dentro del subgrupo	• Error de medición • Error de muestreo • Variación del proceso dentro del ajuste	• Error de medición • Error de muestreo • Variación del proceso dentro del ajuste y entre ajustes
Variación entre subgrupos	• Entre ajustes • Entre días • Entre herramientas • Entre materiales	• Entre días • Entre herramientas • Entre materiales

Tabla 7.9 Componentes de variaciones dentro del subgrupo y entre subgrupos por la formación de subgrupos en las tablas 7.7 y 7.8

(2) Variación dentro del subgrupo y variación entre subgrupos

La variación de los datos se clasifica como variación dentro del subgrupo y variación entre subgrupos. La primera es la variación que aparece al interior de un subgrupo y se encuentra representada por el valor $\overline{R}$ en una gráfica R. Por otra parte, la variación entre subgrupos es la que aparece entre los subgrupos y se encuentra representada por la distribución de los puntos $\overline{x}$ en una gráfica $\overline{x}$.

Sea la varianza al interior de un subgrupo $\sigma_w{}^2$ y la varianza entre subgrupos $\sigma_b{}^2$, entonces

$$\sigma \overline{x}^2 = \sigma_b{}^2 + \frac{\sigma_w{}^2}{n}, \tag{7.1}$$

donde σx^2 es la varianza de $\overline{x}$. σ_w se estima a partir de $\overline{R}$ por

$$\hat{\sigma}_w = \overline{R} / d_2, \tag{7.2}$$

donde d_2 es un coeficiente que depende del tamaño del subgrupo, n, y que se muestra en la tabla 7.4. Aquí el índice "^" en el símbolo σ_ω significa la estimación de σ_ω.

Según esto, la distribución de los puntos $\overline{x}$ se ve afectada no solamente por la variación entre subgrupos sino también por la variación dentro de los subgrupos. La variación de un estado perfectamente controlado es la siguiente.

a) El proceso promedio es constante y la variación entre subgrupos es $\sigma_b{}^2 = 0$.
b) La variación del proceso es constante, es decir, la variación dentro de los subgrupos σw^2 = constante.

No es posible determinar directamente la variación entre subgrupos, sino que se calcula de

$$\hat{\sigma}_b = \sqrt{\hat{\sigma}_{\overline{x}}{}^2 - \hat{\sigma}_w{}^2 / n}, \tag{7.3}$$



donde $\sigma_x^{\ 2}$ se calcula a partir del histograma de $\bar{x}$, o a partir de

$$\hat{\sigma}_b = \sqrt{\hat{\sigma}_x^{\ 2} - \hat{\sigma}_w^{\ 2}}\ , \qquad (7.4)$$

donde $\sigma_x^{\ 2}$ se calcula a partir del histograma de x.

(3) Estratificación

Cuando los mismos productos se hacen en varias máquinas o por varios operarios, es mejor clasificar los datos según la máquina o el operario de manera que la diferencia entre máquinas u operarios pueda ser analizada y el control del proceso se facilite.

La estratificación es un método para identificar la fuente de la variación de los datos recogidos, clasificando los datos según varios factores.

La figura 7.5 es un ejemplo de una gráfica de control que puede estratificarse. Esta gráfica muestra una característica de calidad de partes manufacturadas por dos máquinas (A y B). La

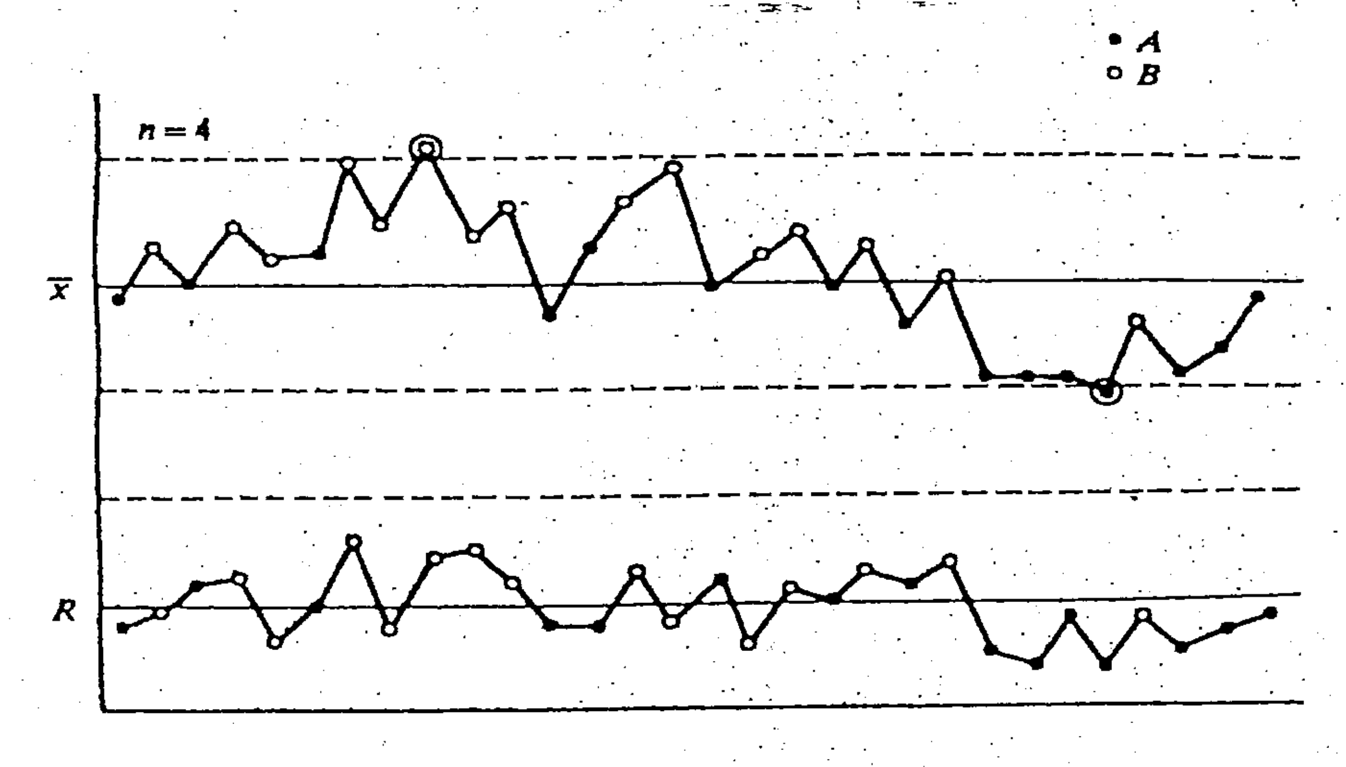


Figura 7.5 Gráfica de control para A y B

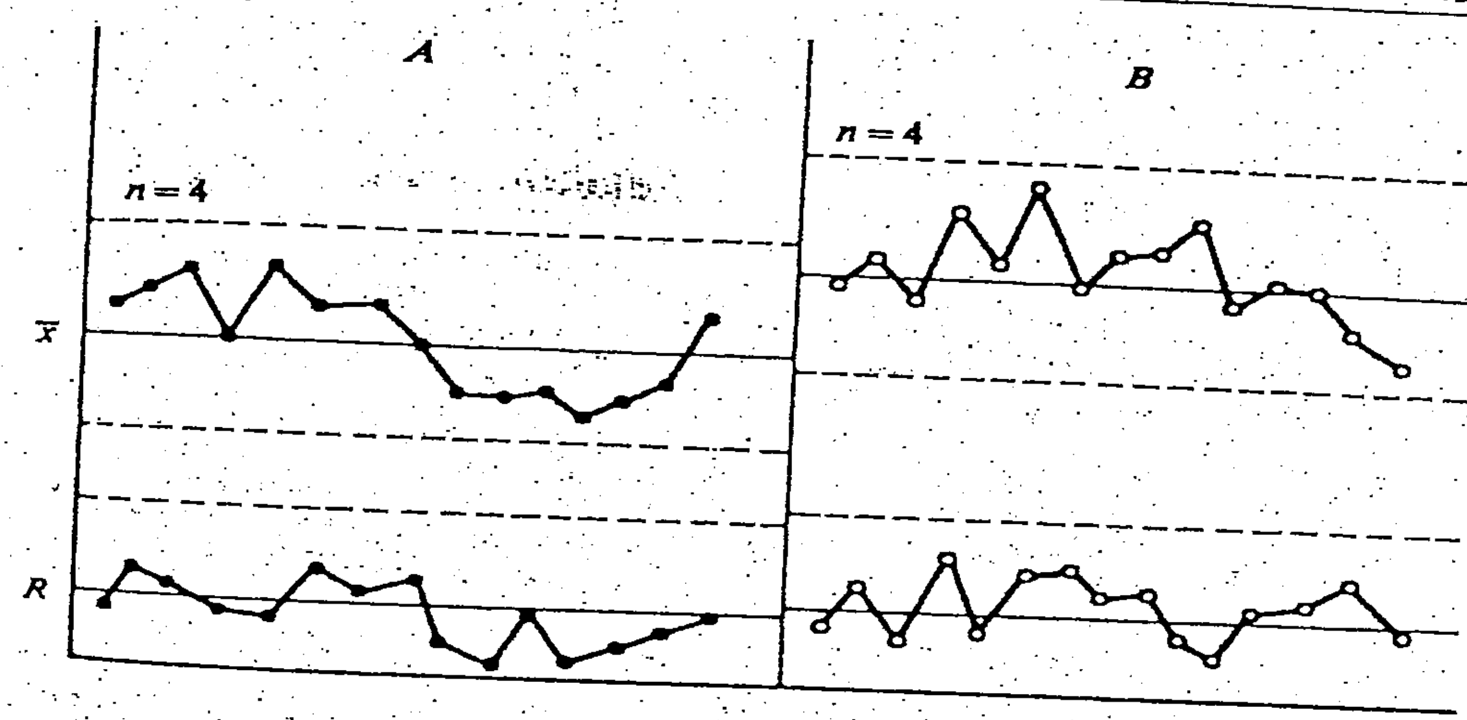


Figura 7.6 Gráficas de control estratificadas para *A* y *B*

figura 7.6 es la gráfica de control estratificada para *A* y *B*. Como resultado de la estratificación se encuentra que casi no hay diferencia en la variación entre *A* y *B*, pero que *B* tiene un valor promedio del proceso mayor que *A* y que el proceso está en estado de control para ambas máquinas, y puede decirse que entre *A* y *B* parece haber algunas causas asignables.

Como regla general, el propósito de la estratificación es examinar la diferencia entre los valores promedio y la variación entre clases diferentes y tomar acciones correctivas con respecto a la diferencia, si la hay. Si es imposible tomar acciones instantáneamente, se hace necesario entonces llevar a cabo el control del proceso usando gráficas de control estratificadas.

(4) Prueba de la diferencia entre gráficas de control estratificadas

Si las dos gráficas estratificadas $\overline{x} - R$ satisfacen las siguientes cuatro condiciones, entonces es posible someter a prueba la diferencia entre los valores promedio:

a) Ambas gráficas muestran un estado de control.
b) El tamaño de los subgrupos es el mismo.
c) Los valores de $\overline{R}_A$ y de $\overline{R}_B$ son casi iguales.
d) Los números de los subgrupos k_A y k_B son suficientemente grandes ($k_A > 10$, $k_B > 10$).

$$\left|\overline{\overline{x}}_A - \overline{\overline{x}}_B\right| \geq A_2 \overline{R} \sqrt{1/k_A + 1/k_B}\,, \tag{7.5}$$

donde

$$\overline{R} = (k_A \overline{R}_A + k_B \overline{R}_B)/(k_A + k_B). \tag{7.6}$$

(5) Prueba de la diferencia en la variación entre estratos

Para ver si hay una diferencia en la variación entre los estratos después de la estratificación, se usa la fórmula (7.7) o (7.8). Si la fórmula resiste, entonces puede decirse que existe una diferencia en la variación entre A y B.

En el caso donde $\overline{R}_A > \overline{R}_B$

$$\overline{R}_A / \overline{R} \geq 1.2 \quad \overline{R} / \overline{R}_B \geq 1.2 \tag{7.7}$$

En el caso donde $\overline{R}_B > \overline{R}_A$

$$\overline{R}_B / \overline{R} \geq 1.2 \quad \overline{R} / \overline{R}_A \geq 1.2 \tag{7.8}$$

Las pruebas mencionadas en (7.5) y (7.7) o (7.8) no son necesarias cuando la diferencia es obvia por la simple inspección de las gráficas de control.

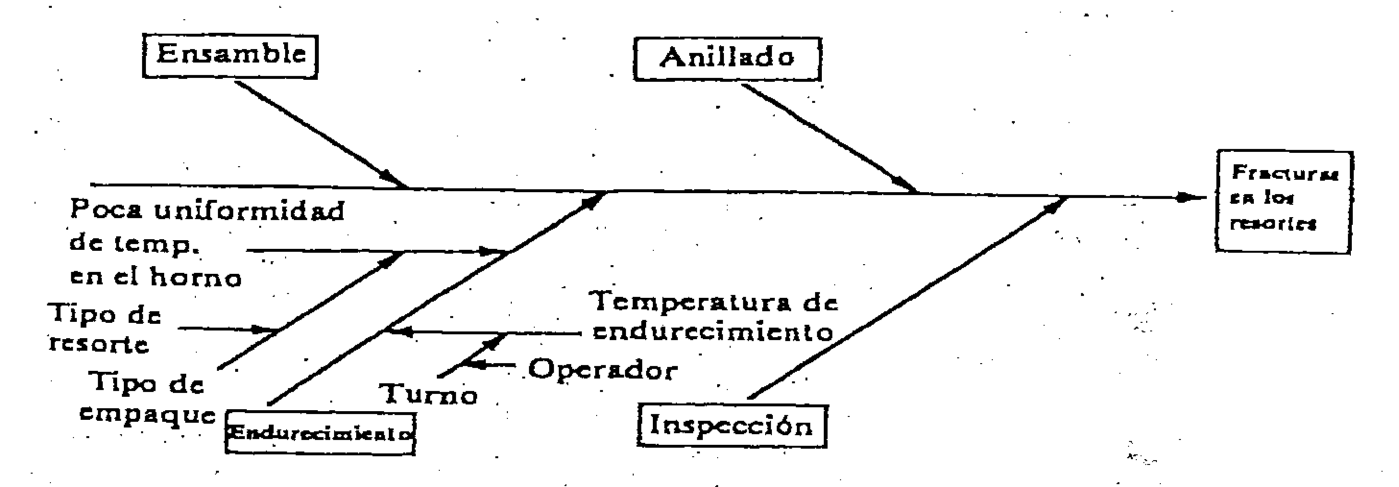


Figura 7.7 Diagrama de causa-efecto de las fracturas en el resorte

7.6 ESTUDIO DE CASO DE ANÁLISIS DEL PROCESO

Por lo general, cuando se usa una gráfica de control para analizar un problema, ésta no se usa sola sino junto con histogramas. A continuación damos un ejemplo de cómo analizar un problema.

Ejemplo 7.2

En una fábrica que producía resortes de hoja para tractores se encontró que algunos de los resortes tenían fracturas. Es necesario identificar la causa del problema tan pronto como sea posible y evitar que vuelva a ocurrir. Usemos los siguientes datos para construir histogramas y gráficas de control para analizar el problema.

Información obtenida hasta el momento

1) Se hizo el anterior diagrama de causa-efecto para identificar la causa de las fracturas.
2) Los resortes para tractores pequeños (A_1) y para tractores medianos (A_2) se tratan en el mismo horno. Los dos tipos de producto se diferencian sólo por la forma, pues el material con el cual se fabrican es el mismo Sin embargo, el tipo de empaque usado durante el tratamiento por calor es diferente.



3) Recientemente ha aumentado la demanda por tractores pequeños, y la producción se está incrementando gradualmente. El número de resortes que reciben tratamiento de calor simultáneamente se ha aumentado proporcionalmente
4) El horno se opera en dos turnos diarios (B_1 y B_2). En cada turno se realizan dos tratamientos de calor, es decir, en cada turno se tratan al calor dos conjuntos de resortes.
5) La dureza de los resortes puede considerarse como una característica sustitutiva de las fracturas.
6) Durante los últimos 16 días se ha recogido la siguiente información:

   a) Se pensó que la variación en la temperatura del horno del centro hacia las paredes podría ser significativa. Por tanto, después de cada tratamiento se tomó una muestra (P_1) del centro del horno y otra (P_2) de cerca a la pared, y se midió su dureza.
   b) Se inspeccionaron todos los resortes después del tratamiento de calor y se midió su dureza.
   c) Los estándares de dureza son los siguientes: dureza máxima ~~460 Hb,~~ dureza mínima 350 Hb.

Información obtenida de los histogramas

1) Histograma general (figura 7.8)

Éste muestra aproximadamente una distribución normal y todas las muestras se encuentran dentro del rango especificado de dureza. Sin embargo, se encontraron fracturas en algunas muestras de alta dureza aunque estaban dentro de los límites de la especificación

2) Histograma para las muestras con fracturas (figura 7.9)

a) Todas las muestras con fracturas son del tipo A_1. Por tanto, el problema se localiza en el tratamiento de resortes tipo A_1 para tractores pequeños.

Día	Tipo de resorte A	Turno B	No. Lote	Posición P_1	Posición P_2	Dureza resorte con fracturas
1	A_1	B_1	1	396	420	[illegible]
			2	396	421	
		B_2	3	408	423	460
			4	408	438	
2	A_1	B_1	5	393	400	
			6	401	399	
		B_2	7	404	438	
			8	396	429	450
3	A_1	B_1	9	385	410	451
			10	391	432	456 453
		B_2	11	377	407	
			12	378	410	
4	A_1	B_1	13	387	421	456 443
			14	397	422	
		B_2	15	397	397	462 446 455
			16	384	404	
5	A_2	B_1	17	402	391	
			18	398	401	
		B_2	19	393	382	
			20	381	366	
6	A_2	B_1	21	392	411	
			22	382	399	
		B_2	23	395	402	
			24	407	381	
7	A_2	B_1	25	413	392	
			26	387	392	
		B_2	27	394	409	
			28	401	409	
8	A_2	B_1	29	401	404	
			30	400	404	
		B_2	31	414	418	
			32	406	407	
9	A_1	B_1	33	406	418	453 457
			34	397	421	
		B_2	35	436	419	
			36	400	454*	454* 449
10	A_1	B_1	37	390	432	
			38	387	422	450
		B_2	39	398	409	
			40	378	419	
11	A_1	B_1	41	390	420	
			42	417	430	445 458 473 446
		B_2	43	373	419	457 455 465
			44	385	395	458
12	A_1	B_1	45	394	406	460 455
			46	391	410	
		B_2	47	385	413	
			48	378	419	447 444 457
13	A_2	B_1	49	411	403	
			50	410	392	
		B_2	51	385	370	
			52	398	393	
14	A_2	B_1	53	394	395	
			54	397	419	
		B_2	55	409	406	
			56	397	404	
15	A_2	B_1	57	406	399	
			58	411	415	
		B_2	59	385	386	
			60	408	418	
16	A_2	B_1	61	387	410	
			62	395	401	
		B_2	63	410	395	
			64	400	409	

Tabla 7.10 Mediciones de la dureza

* Los asteriscos indican el mismo resorte, pues se encontraron fracturas en la muestra P_2.

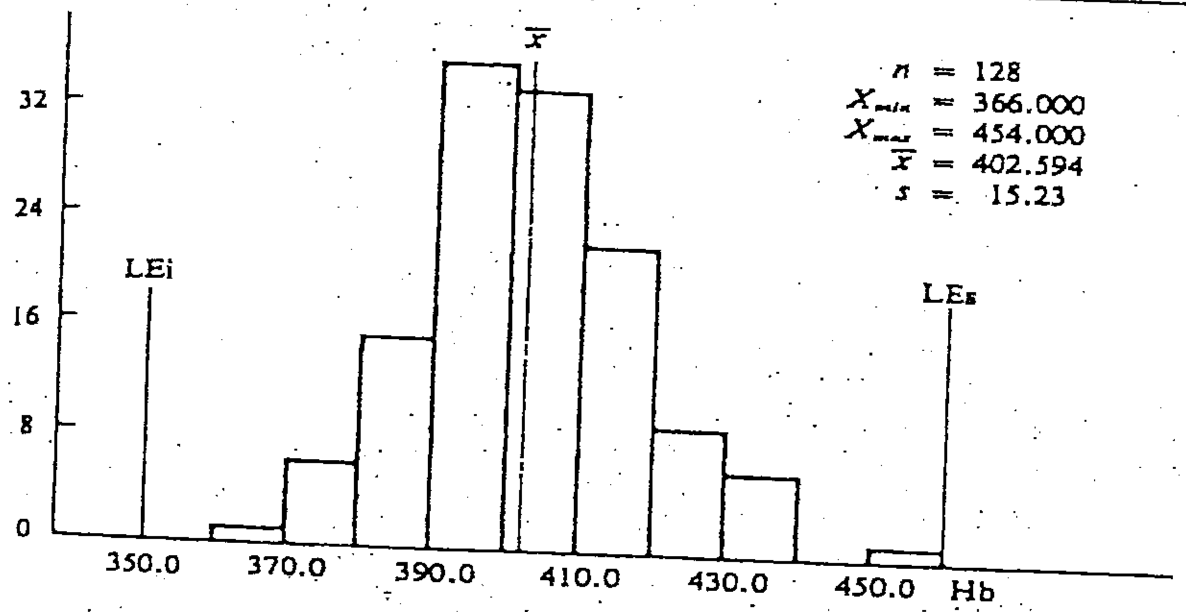


Figura 7.8 Histograma general

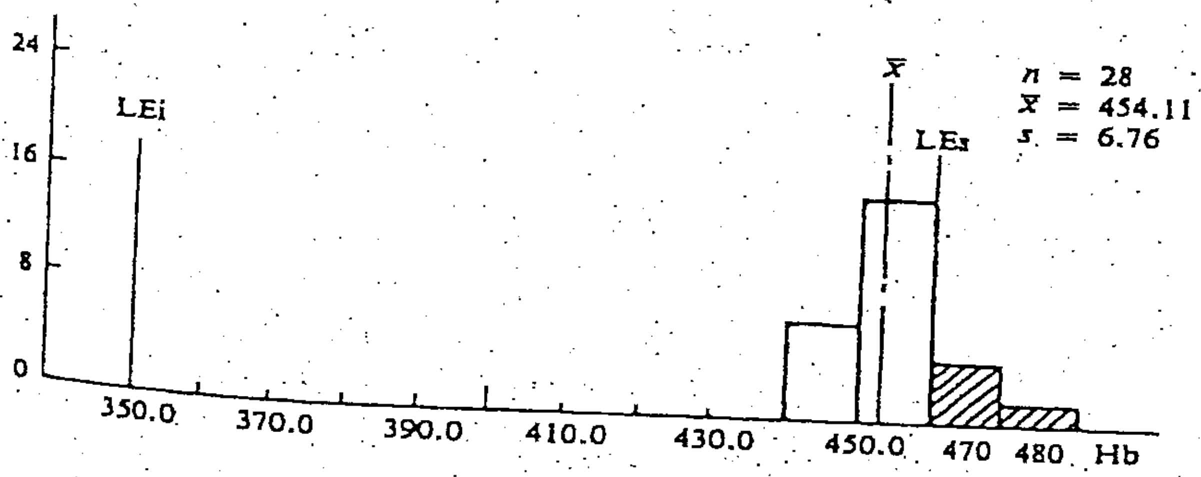


Figura 7.9 Histograma de muestras con fracturas

b) La dureza de los resortes con fracturas se distribuye alrededor de un valor alto y todos son más altos que 440 Hb.

3) Histogramas para diferentes tipos de resorte A_1 y A_2 (figura 7.10)

a) La dureza promedio de los resortes para los tractores pequeños (A_1) es más bien alta, y los valores individuales se distribuyen más ampliamente.

b) Debido a que todos los resortes con fracturas son de muestras del tipo A_1, parecería que los métodos de producción para los resortes para tractores pequeños no son adecuados.



c) La distribución de los valores de dureza para los resortes tipo A_2 para tractores medianos no es amplia, y no se encontraron entre ellos muestras con fracturas.

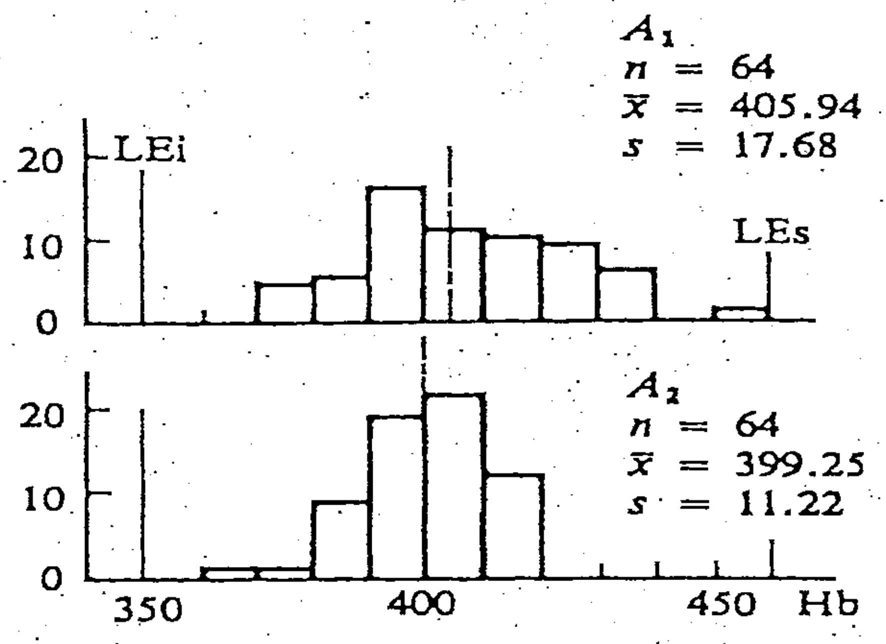


Figura 7.10 Histogramas para diferentes tipos de resortes A_1 y A_2

4) Histogramas para diferentes turnos B_1 y B_2 (figura 7.11)

a) La medida para B_1 es mayor que la de B_2 y la variación es menor.

b) Entre las muestras B_2 hay algunos resortes con fracturas.

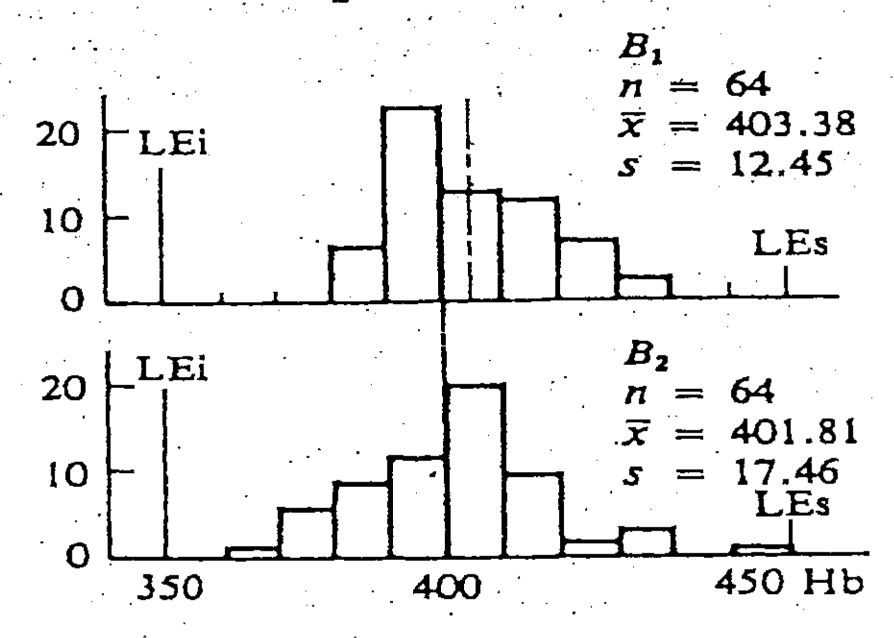


Figura 7.11 Histogramas para diferentes turnos B_1 y B_2

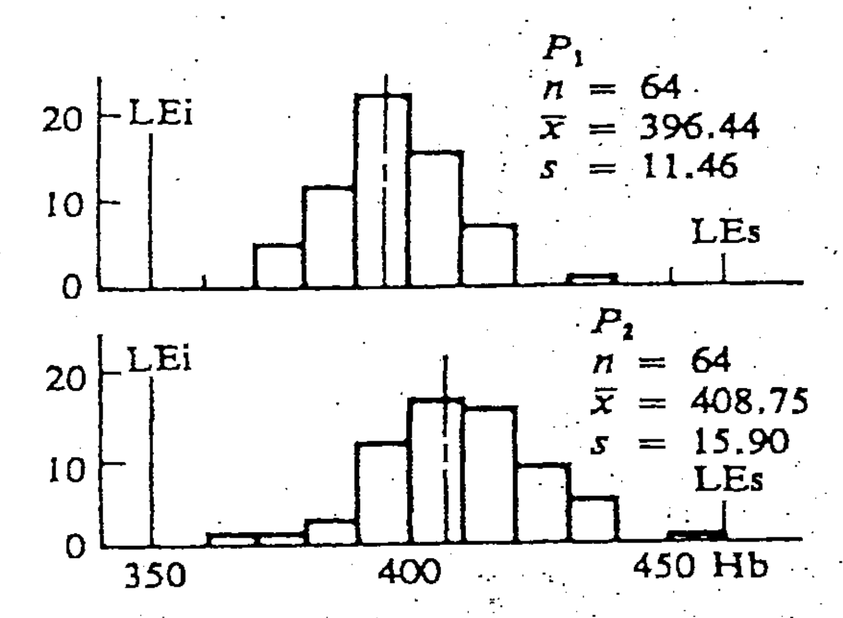


Figura 7.12 Histogramas para las diferentes posiciones P_1 y P_2

5) Histogramas para diferentes posiciones en el tratamiento de calor P_1 y P_2 (figura 7.12)

a) La dureza media de las muestras tomadas del centro del horno (P_1) es baja, y la variación es pequeña. Ninguna de estas muestras tiene fracturas.

b) La dureza media de las muestras tomadas de cerca a la pared del horno (P_2) es alta y la variación es ligeramente mayor. Algunas de estas muestras tienen fracturas.

c) Parece que aquellos resortes tratados cerca a la pared del horno tienen probabilidad de tener una dureza lo suficientemente grande como para producir fracturas.

6) Histogramas para diferentes combinaciones de factores A y B (figura 7.13)

a) La variación para las combinaciones $A_1\ B_1$ y $A_1\ B_2$ es mayor que para $A_2\ B_1$ o para $A_2\ B_2$.

b) No hay diferencia significativa en las medias, pero aquéllas para $A_1\ B_1$ y $A_1 B_2$ son ligeramente mayores.

c) Los valores para la combinación $A_2 B_1$ se concentran alrededor de la media de la dureza especificada, y la variación es pequeña. Basado en la actual dureza máxima y mínima, el



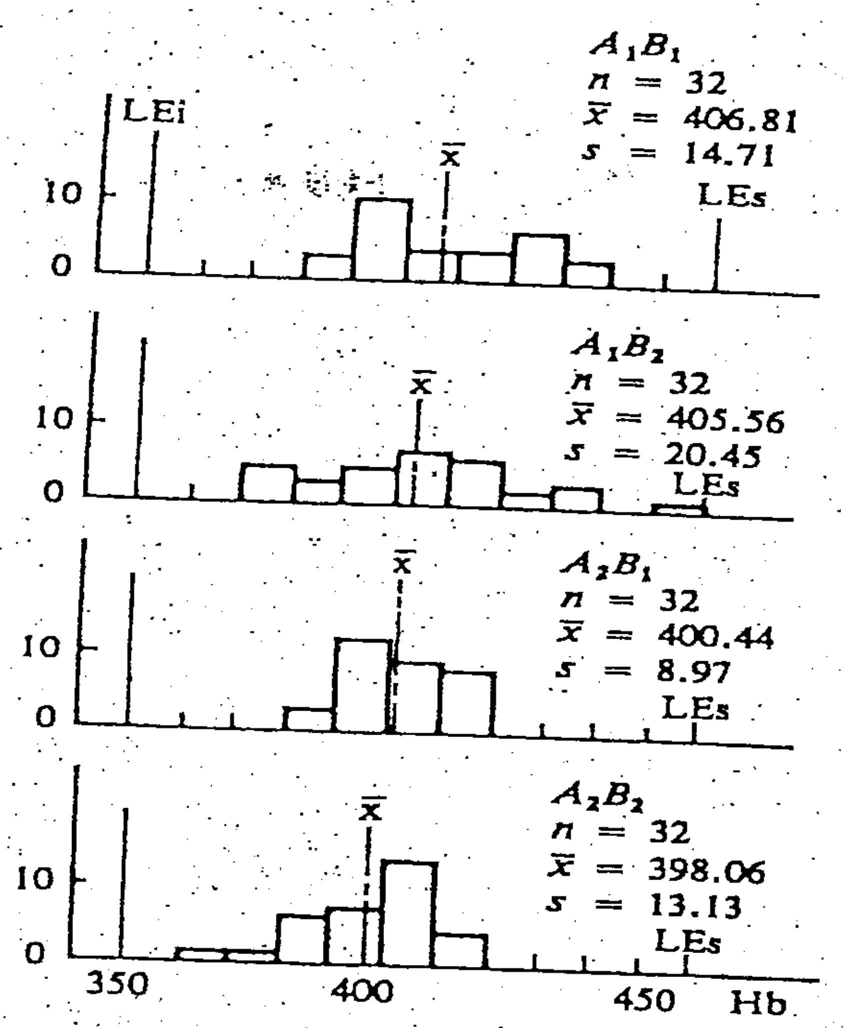


Figura 7.13 Histogramas para diferentes combinaciones de dos factores A y B

índice de capacidad del proceso Cp tiene un valor extremadamente bueno de 2.04, calculado por

$$C_P = \frac{\text{LEs}}{6s} - \text{LEi} = \frac{460-350}{6\times 8.97} = 2.04.$$

Por tanto, la combinación de A_2B_1 es excelente.

7) *Histogramas para diferentes combinaciones de factores A, B y P (figura 7.14)*

a) La dureza media para las muestras de tipo A_1 (para tractores pequeños) es claramente mayor para las muestras tomadas de la posición P_2 (cerca a la pared del horno) que para las muestras tomadas de la posición P_1 (el centro del horno).

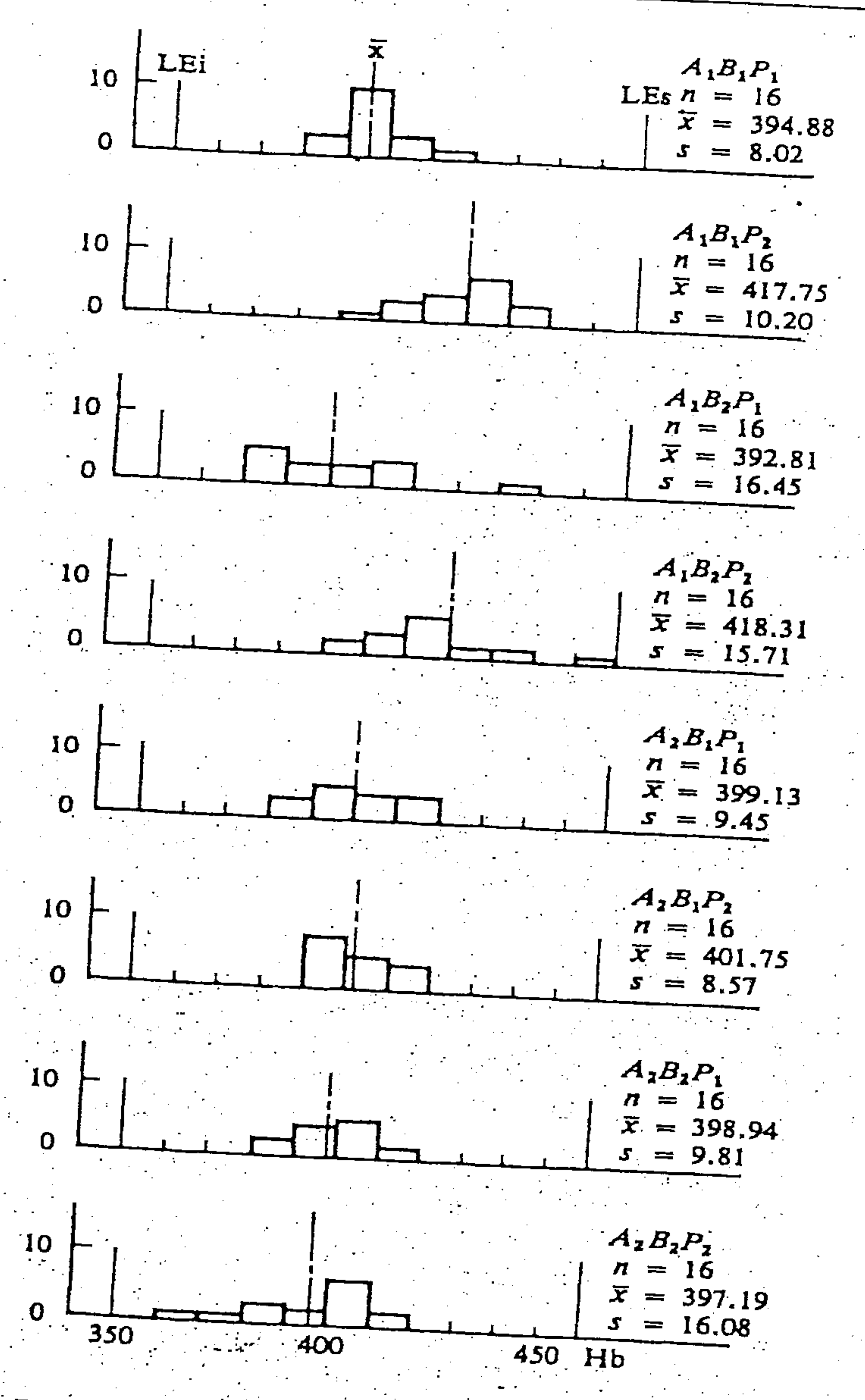


Figura 7.14 Histogramas para diferentes combinaciones de tres factores A, B y P

b) La dureza media para las muestras tipo A_2 (para tractores medianos) no parece depender de la posición de la cual se tomen las muestras.

c) La desviación estándar tiene el mismo valor de alrededor de 10 Hb para las combinaciones $(A_1B_1P_1)$, $(A_1B_1P_2)$, $(A_2B_1P_1)$, $(A_2B_1P_2)$ y $(A_2B_2P_1)$. Tiene un valor mayor de alrededor de 15 Hb para las combinaciones $(A_1B_2P_1)$, $(A_1B_2P_2)$ y $(A_2B_2P_2)$. Esta diferencia parece deberse al factor B, debiéndose la mayor variación a B_2.

d) Las combinaciones $(A_2B_1P_1)$, $(A_2B_1P_2)$ y $(A_2B_2P_1)$ se concentran alrededor de la media de la dureza especificada y la variación es pequeña.

Información obtenida de la gráfica $\overline{x} - R$ (figura 7.15)

1) Gráfica de control general

Debido a que los datos están divididos en subgrupos por lotes, la variación en el interior de un subgrupo es la variación dentro de un lote, es decir, la variación entre P_1 y P_2, el error de muestreo y el error de medición. La variación entre los diferentes subgrupos incluye la variación entre lotes y entre turnos, diferentes tipos de resortes y diferentes días (tabla 7.11).

a) Gráfica R

i) Ninguno de los valores está por fuera de los límites de control, pero las largas series de valores relativa y suavemente conectados para los lotes 7-14, 26-33 y 51-64 son anormales, e indican que el proceso está fuera de control.

ii) R es relativamente grande para resortes del tipo A_1 (tractor pequeño) y relativamente pequeño para resortes del tipo A_2 (tractores medianos). Estos dos tipos, por tanto, deben clasificarse por separado.

b) Gráfica $\overline{x}$

No hay puntos por fuera de los límites de control, pero la larga serie de valores para los lotes 15-26, 31-39 y 43-53 son anormales, indicando un estado fuera de control.



Día	Tipo de resorte A	Turno B	Lote No.	Posición	
				P_1	P_2
1	A_1	B_1	1	396	420
			2	396	421
		B_2	3	408	423
			4	408	438
2	A_1	B_1	5	393	400
			6	401	399
		B_2	7	404	438
			8	396	429
3	A_1	B_1	9	385	410
			10	391	432
		B_2	11	377	407
			12	378	410

Tabla 7.11 Subgrupos de la gráfica de control general

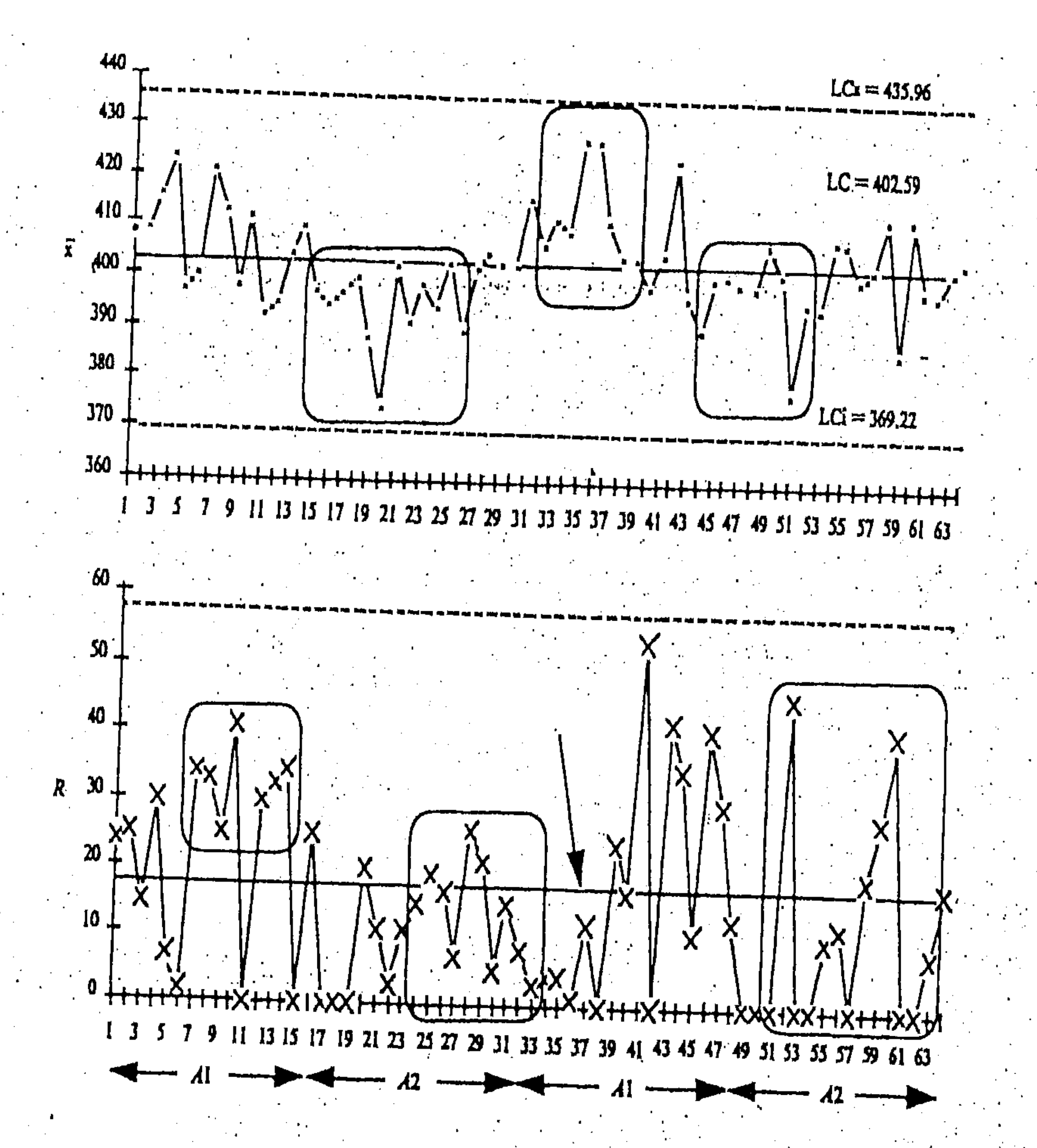


Figura 7.15 Gráfica general $\bar{x}$ – R

2) *Gráficas de control estratificadas según factores A y B (figura 7.16)*

Debido a que los datos están estratificados por lote, la variación en el interior de un subgrupo es la misma que la variación en el interior del lote correspondiente. La variación entre los subgrupos incluye la variación entre lotes y la variación diaria (tabla 7.12).

Día	Tipo de resorte A	Turno B	Lote No.	Posición P_1	Posición P_2
1	A_1	B_1	1	396	420
			2	396	421
		B_2	3	408	423
			4	408	438
2	A_1	B_1	5	393	400
			6	401	399
		B_2	7	404	438
			8	396	429
3	A_1	B_1	9	385	410
			10	391	432
		B_2	11	377	407
			12	378	410

Tabla 7.12 Subgrupos de gráficas de control estratificadas según dos factores *A* y *B*

a) (A_1B_1): Tanto la gráfica R como la gráfica $\bar{x}$ muestran una aproximación a la línea central. Cada subgrupo se compone de datos sobre resortes tomados tanto del centro del horno como de cerca a la pared. Esta tendencia es, por tanto, el resultado de las diferentes medias de las dos distribuciones. Es necesario estratificar los datos según la posición (P).

b) (A_1B_2): Puede decirse lo mismo que sobre las gráficas para (A_1B_1).

c) (A_2B_1): Tanto la gráfica R como la $\bar{x}$ están en estado controlado. En otras palabras, no hay variación entre los subgrupos (es decir, entre los lotes y día a día). El proceso está en operación estable. Estimemos la desviación estándar en el interior del lote, σ_ω.

En este caso

$$n = 2, \; d_2 = 1.128,$$

tenemos

$$\hat{\sigma}_w = \bar{R} / d_2 = 10.75/1.128 = 9.53$$

Esto es aproximadamente el mismo valor de σ_x donde $\sigma_x = s = 8.97$, estimado del histograma de la figura 7.13 (A_2B_1).

d) (A_2B_2): La gráfica R está en estado de control. Sin embargo, la gráfica $\bar{x}$ incluye tres puntos fuera de los límites de control, y no está en estado de control. Hay una variación entre los subgrupos; es decir, entre lotes diferentes y días diferentes. Debe examinarse el método usado en el turno B_2. La variación estimada entre lotes, σ_ω, es

$$\hat{\sigma}_w = \bar{R} / d_2 = 8.42$$

La desviación estándar obtenida de la figura 7.13 (A_2B_2) es 13.13.

Puesto que $\sigma_x^{\;2}$ se expresa por

$$\sigma_x^{\;2} = \sigma_w^{\;2} + \sigma_b^{\;2} ,$$

la variación entre lotes σ_b está dada por

$$\hat{\sigma}_b = \sqrt{\hat{\sigma}_x^{\;2} - \hat{\sigma}_w^{\;2}} = \sqrt{13.13^2 - 8.42^2} = 10.07 .$$

En otras palabras, la variación entre subgrupos es del mismo orden que la variación en el interior de subgrupos.



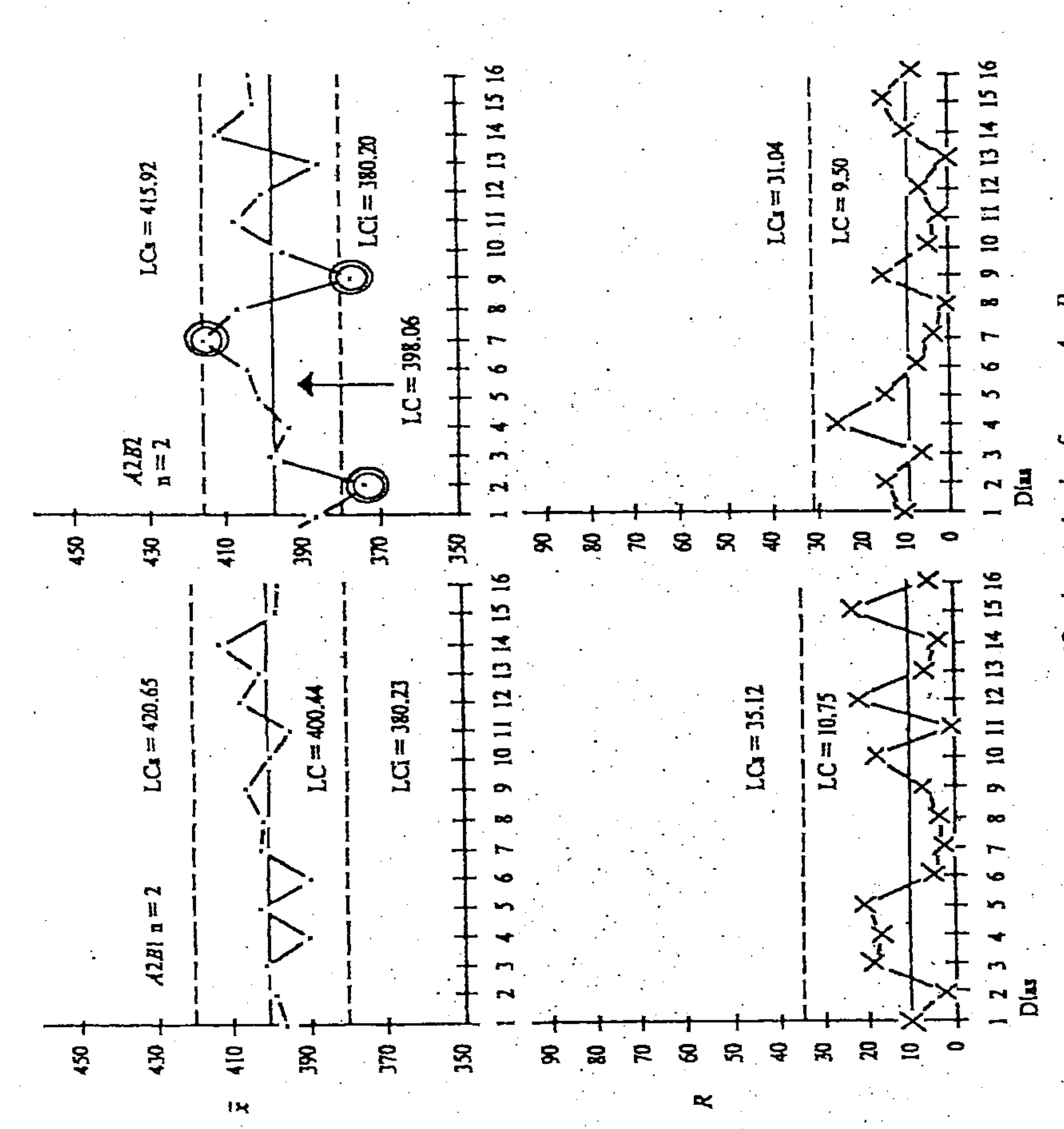


Figura 7.16 Gráficas $\bar{x}$ – R estratificadas según dos factores A y B

3) *Gráficas de control para resortes de tipo A_1 estratificadas según B y P (figura 7.17)*

Debido a que la media difiere según la posición en el horno, se hicieron gráficas de control según posición (P), combinando los datos de dos lotes cada día en un subgrupo. Así, la variación en el interior de un subgrupo consiste en la variación durante un solo día, la cual incluye la variación entre lotes. La variación entre subgrupos consiste en la variación entre días (tabla 7.13).

Día	Tipo de resorte A	Turno B	Lote No.	Posición	
				P_1	P_2
1	A_1	B_1	1	396	420
			2	396	421
		B_2	3	408	423
			4	408	438
2	A_1	B_1	5	393	400
			6	401	399
		B_2	7	404	438
			8	396	429
3	A_1	B_1	9	385	410
			10	391	432
		B_2	11	377	407
			12	378	410

Tabla 7.13 Subgrupos de gráficas de control para resortes tipo A_1 estratificadas según B y P

a) La diferencia en $\bar{x}$ para $(A_1B_1P_1)$ y $(A_1B_1P_2)$ es obvia. Un punto se encuentra por fuera de los límites de control, pero tanto la gráfica $\bar{x}$ como la gráfica R muestran un estado de control relativamente estable.

b) Una comparación de P_1 y P_2 para la combinación A_1B_1. Puesto que $\overline{R}$ es aproximadamente la misma para $(A_1B_1P_1)$ y $(A_1B_1P_2)$, los datos para éstos pueden reunirse para estimar σ_ω, entonces

$$\hat{\sigma}_w = \overline{R} / d_2 = \frac{8.25 + 6.50}{2 \times 1.128} = 6.54 .$$

Respecto a σ_b que incluye la diferencia dependiendo de la posición, ésta se obtiene por

$$\hat{\sigma}_b = \sqrt{\hat{\sigma}_x^{\ 2} - \hat{\sigma}_w^{\ 2}} = \sqrt{14.71^2 - 6.54^2} = 13.18 .$$

La contribución de la variación entre subgrupos a la variación total $\sigma_x^{\ 2}$ se puede expresar como un porcentaje de ρ_b.

$$\hat{\rho}_b = \hat{\sigma}_b^{\ 2} / \hat{\sigma}_x^{\ 2} \times 100 = (13.18^2 / 14.71^2) \times 100 = 80\% .$$

En este caso, $\sigma_b^{\ 2}$ depende de la posición en el horno, indicando también que la dureza de los resortes está sesgada según su posición en el horno, y que esto debe mejorarse.

c) Si comparamos la figura 7.16 (A_1B_1) con la figura 7.17 $(A_1B_1P_1)$, veremos que en la última gráfica, la estratificación de la primera por P ha dado como resultado un estado de control estable. Esto se debe a que se ha cambiado la composición de los subgrupos. Los datos deben agruparse de manera que la variación dentro del subgrupo sea lo más aleatoria posible, haciendo que la variación debida al proceso aparezca como variación entre subgrupos.

A1B1 n = 2

LCs = 449.79

LC = 406.31

LCi = 362.84

LCs = 75.55

LC = 23.13

A1B2 n = 2

LCs = 457.50

LC = 405.56

LCi = 353.63

LCs = 90.25

LC = 27.63

Tabla 7.17 Gráficas $\bar{x} - R$ para resortes tipo A_1 estratificadas según B y P



4) *Gráficas de control para resortes tipo A_2 estratificadas según B (figura 7.18)*

Respecto al tipo A_2 no hay diferencia en la dureza según la posición en el horno, y los datos se agruparon en dos lotes cada día con $n = 4$. De esta manera, la variación intra-subgrupos se compone de la variación diaria, incluyendo tanto la variación en el interior de los lotes como entre ellos. La variación inter-subgrupos es la variación entre días.

a) (A_2B_1): Las gráficas $\bar{x}$ y R están ambas en el estado de control. $\sigma_{(t)}$ se estima a partir de

$$\hat{\sigma}_{w} = \bar{R} / d_2 = 19.13 / 2.059 = 9.29 ,$$

b) (A_2B_2): La gráfica R está en estado de control, pero la gráfica $\bar{x}$ no lo está. Un punto se encuentra por fuera de los límites de control y se observa un acercamiento de los límites de control.

Día	Tipo de resorte A	Turno B	Lote Nº	Posición P_1	Posición P_2
4	A_2	B_1	13	387	421
			14	397	422
		B_2	15	397	397
			16	384	404
5	A_2	B_1	17	402	391
			18	398	401
		B_2	19	393	382
			20	381	366
6	A_2	B_1	21	392	411
			22	382	399
		B_2	23	395	402
			24	407	381

Tabla 7.14 Subgrupos de gráficas de control para resortes tipo A_2 estratificadas según B



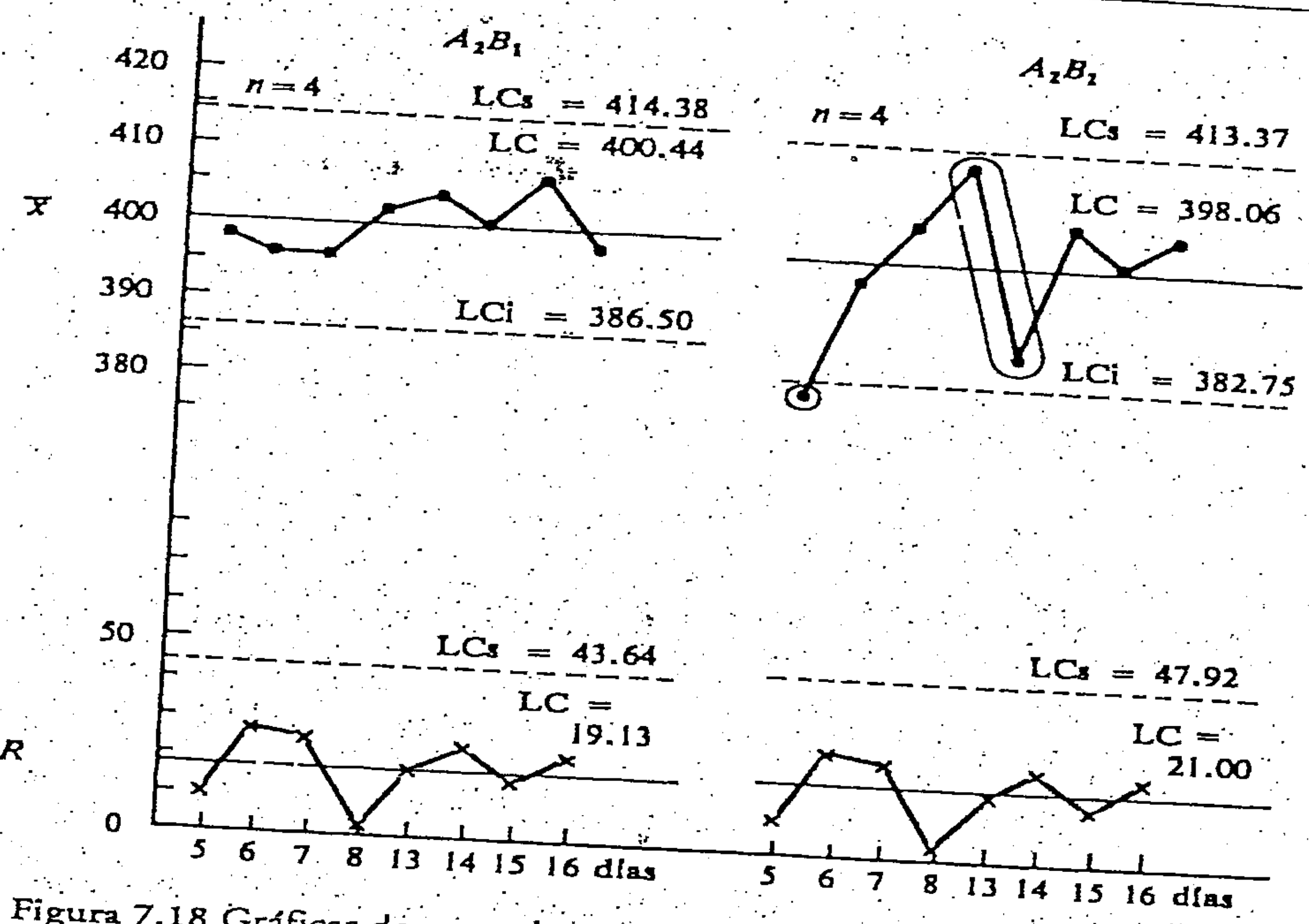


Figura 7.18 Gráficas de control para resortes tipo A_2 estratificadas según B

Resumen y medidas remediales

1) Las fracturas han aparecido en los resortes para los tractores pequeños. Se cree que esto se debe al método por el cual se empacan en el horno de tratamiento los resortes para tractores pequeños. Cuando los resortes pequeños reciben el tratamiento al calor se produce una diferencia en la temperatura en el centro del horno en comparación con la posición cerca a la pared. Se cree que esto aumenta la dureza de los resortes y da lugar a fracturas. Es necesario, por tanto, cambiar el método de empaque por uno que no cause una diferencia de temperatura mayor que la que actualmente se produce cuando se tratan los resortes de tamaño mediano.
2) Hay un problema con la fijación de los estándares de operación. Es necesario establecer la relación entre la dureza de los

resortes y su tendencia a fracturarse, y fijar nuevos estándares. Como una estimación temporal basada en los datos dados aquí, se debe fijar la dureza máxima en 440 Hb, por debajo de la cual no se encuentran fracturas. Si se controla la dureza por debajo de 440 Hb, pueden no ocurrir fracturas.

3) Una comparación entre los diferentes turnos mostró que la operación en el turno B_2 no era estable. Se deben revisar los estándares de operación y los métodos. Se debe investigar la razón por la cual no hay variación entre lotes en un día pero sí la hay de día a día.

4) Si se toman las anteriores medidas, es de esperar que se logre la capacidad de proceso obtenida actualmente con la combinación A_2B_1, tanto para resortes pequeños como para resortes de tamaño mediano, los defectos se eliminarán y el índice de capacidad del proceso será

$$Cp = \frac{\text{LEs} - \text{LEi}}{6s} = \frac{440 - 350}{6 \times 8.97} = 1.67 \, .$$

7.7 CONTROL DEL PROCESO CON GRÁFICAS DE CONTROL

Cuando se ha captado suficientemente la relación entre una característica de calidad y los factores de proceso que la afectan, el paso siguiente es controlar estos factores en ciertos niveles de manera que el valor esperado de la característica de calidad se mantenga dentro de un rango deseable. Este paso se llama control del proceso. La gráfica de control sirve como un medio útil para identificar condiciones anormales de los procesos y para mantener los procesos en condición estable.

(1) Características de control

Una variable que se usa para conducir el control de un proceso se llama la característica de control del proceso. Las siguientes con-



sideraciones deben tenerse en cuenta en la determinación de las características de control.

a) Los valores característicos deben reflejar correctamente los estados del proceso.
b) Deben minimizarse los efectos de áreas externas.
c) Los resultados deben estar disponibles de inmediato.
d) El muestreo y la medición deben ser económicos.

Una característica sustitutiva puede usarse si tiene una relación fuerte con la característica de control original. Para el caso en que la medición implique la destrucción de los materiales, puede hacerse una sustitución por mediciones no destructivas.

(2) Determinación de líneas de control

Para administrar un proceso por medio del uso de una gráfica de control, es necesario examinar si la capacidad del proceso es adecuada; es decir, si el proceso es estable y si los rangos de variación de la característica de control en la gráfica indican conformidad satisfactoria con el estándar requerido para producir cierto producto. Cuando se encuentra que el proceso es inadecuado y su característica de control no está en un estado de control, es necesario iniciar actividades tentativas de control contra las anormalidades fijando líneas temporales de control, y al mismo tiempo mejorar el proceso.

La gráfica de control que se hace para el análisis del proceso se compara con los valores estándar. Si la gráfica de control muestra que el proceso está en el estado deseado, entonces las líneas de control se extienden para el control del proceso. Su procedimiento se da en la figura 7.19.



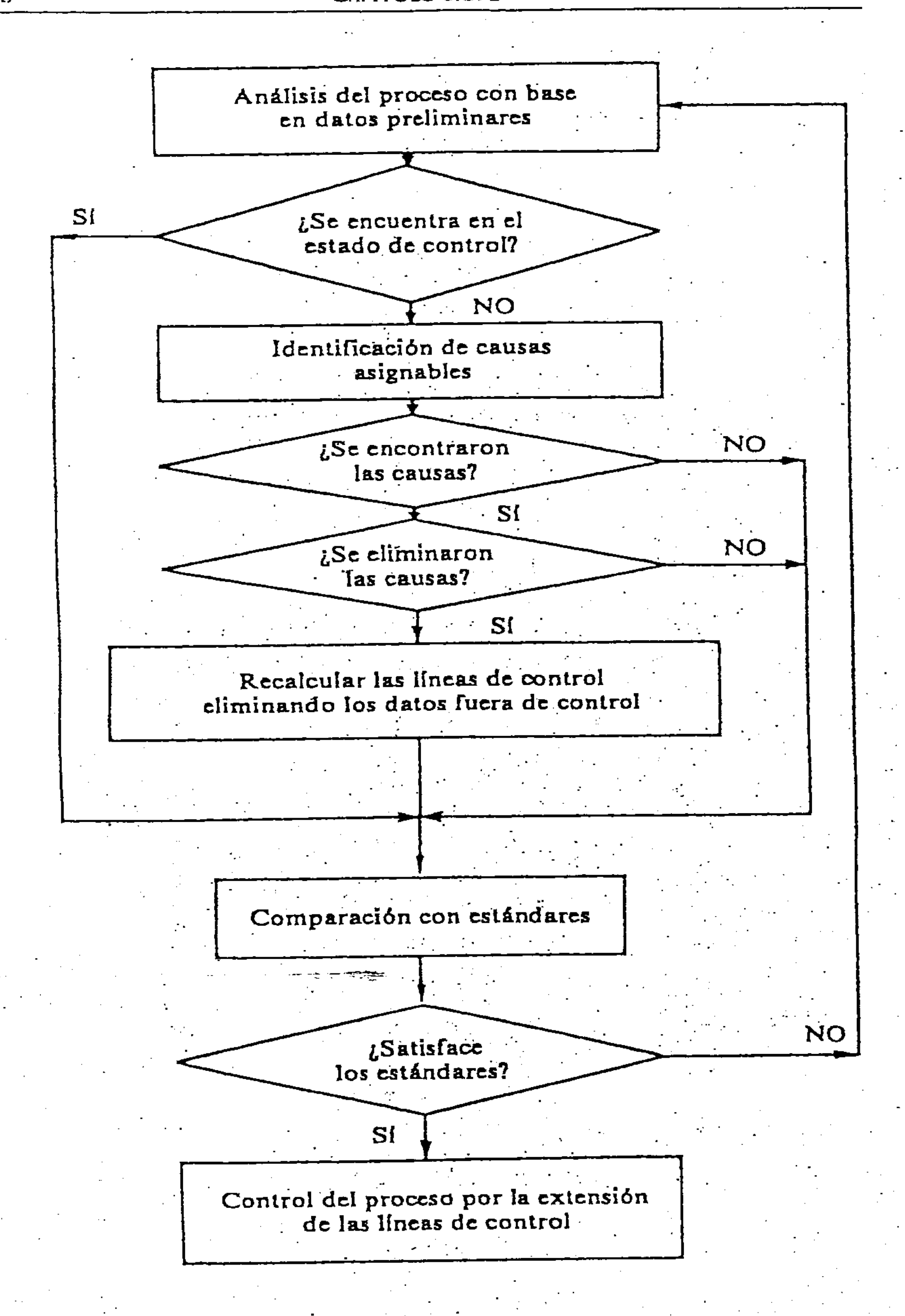


Figura 7.19 Flujo de decisión de líneas de control

(3) Revisión de las líneas de control

Es necesario revisar las líneas de control cuando se reconocen cambios técnicos en el estado del proceso. Esta revisión debe hacerse lo más pronto posible. Aun cuando no se reconocen cambios conspicuos, deben mantenerse revisiones regulares de las líneas de control. Su revisión debe basarse en el rango de las fluctuaciones que pueden ocurrir cuando el proceso está bien controlado.

(4) Estándares de operación

Para poner todo un proceso en un estado estable por medio del control del proceso, es necesario captar los factores que contribuyen a las fluctuaciones del proceso y evitar cambios anormales de estos factores. Para lograrlo se requiere la estandarización de los procedimientos de operación y de los métodos. Para proponer un conjunto de estándares de operación se deben tener en cuenta las siguientes consideraciones:

a) La estandarización debe ser consistente con los objetivos mencionados.
b) Los estándares deben establecerse para controlar la fluctuación de los factores contribuyentes.
c) Los estándares deben ser prácticos y servir como criterio de operación.
d) Son decisiones tentativas y no necesariamente metas ideales.
e) Deben especificar los procedimientos importantes.
f) Se deben hacer revisiones de los estándares para mejorar.
g) Debe comprenderse claramente el contexto de elaboración de los estándares, y debe aclararse el proceso para fijar los estándares.
h) Los estándares deben fijar claramente la responsabilidad y la autorización.
i) En la documentación de estándares debe tenerse en cuenta la facilidad de utilización de los manuales.

j) Se deben describir las medidas temporales para emergencias.
k) Deben tenerse en cuenta las consideraciones a prueba de errores y para la seguridad.
l) Deben estar orientados hacia las metas, no hacia el formalismo.
m) Deben implementarse su instrucción y entrenamiento.

En la planeación de estándares, es indispensable la habilidad para controlar los principales factores del proceso. También, el éxito de la estandarización depende de la devoción de los trabajadores a los estándares. En la realización del proceso de control, los estándares deben revisarse continuamente hasta la perfección, usando la gráfica de control. Además, también se deben establecer los procedimientos relacionados con la estandarización, tales como aplicación, documentación, revisión, educación y entrenamiento.

(5) Comparación con las especificaciones

Cuando haya que comparar los datos del proceso con los límites de especificación, hágalo con la unidad determinada en las especificaciones. Si los límites de la especificación se aplican a cada ítem individual, los límites de la especificación deben compararse con los datos de cada ítem, no con $\bar{x}$, ni con los límites de control.

Si el histograma se encuentra dentro del límite superior e inferior de la especificación con un margen, puede juzgarse satisfactorio para los estándares. Entonces, este proceso puede controlarse por las líneas de control calculadas a partir de los datos del proceso.

Por otra parte, si el histograma rebasa los límites superior e inferior, significa que el proceso no es satisfactorio. Se requieren procedimientos correctivos.

Aunque un proceso se encuentre en el estado de control, el proceso podría producir productos defectuosos, y viceversa. Los límites de control se hacen para proporcionar un juicio sobre si



el proceso está en el estado de control o no, mientras que los límites de especificación se hacen para proporcionar un juicio sobre si cada producto es defectuoso o no. El estado de control es un estado en el cual las causas asignables se eliminan y la variación del proceso se debe sólo al azar. Puede obtenerse realizando el trabajo según las instrucciones de trabajo. La amplitud de los límites de control se determina basándose en la variación debida al azar. Por otra parte, los límites de especificación se deciden por las demandas de los consumidores o los usuarios.

Así, como se muestra en la figura 7.20, es probable que se presenten los siguientes cuatro casos:

a) El proceso no está en un estado de control y también se producen ítems defectuosos.
b) El proceso está en estado de control y sin embargo se producen ítems defectuosos.
c) El proceso no está en estado de control pero los productos no son defectuosos.
d) El proceso está en estado de control pero los productos tampoco son defectuosos.

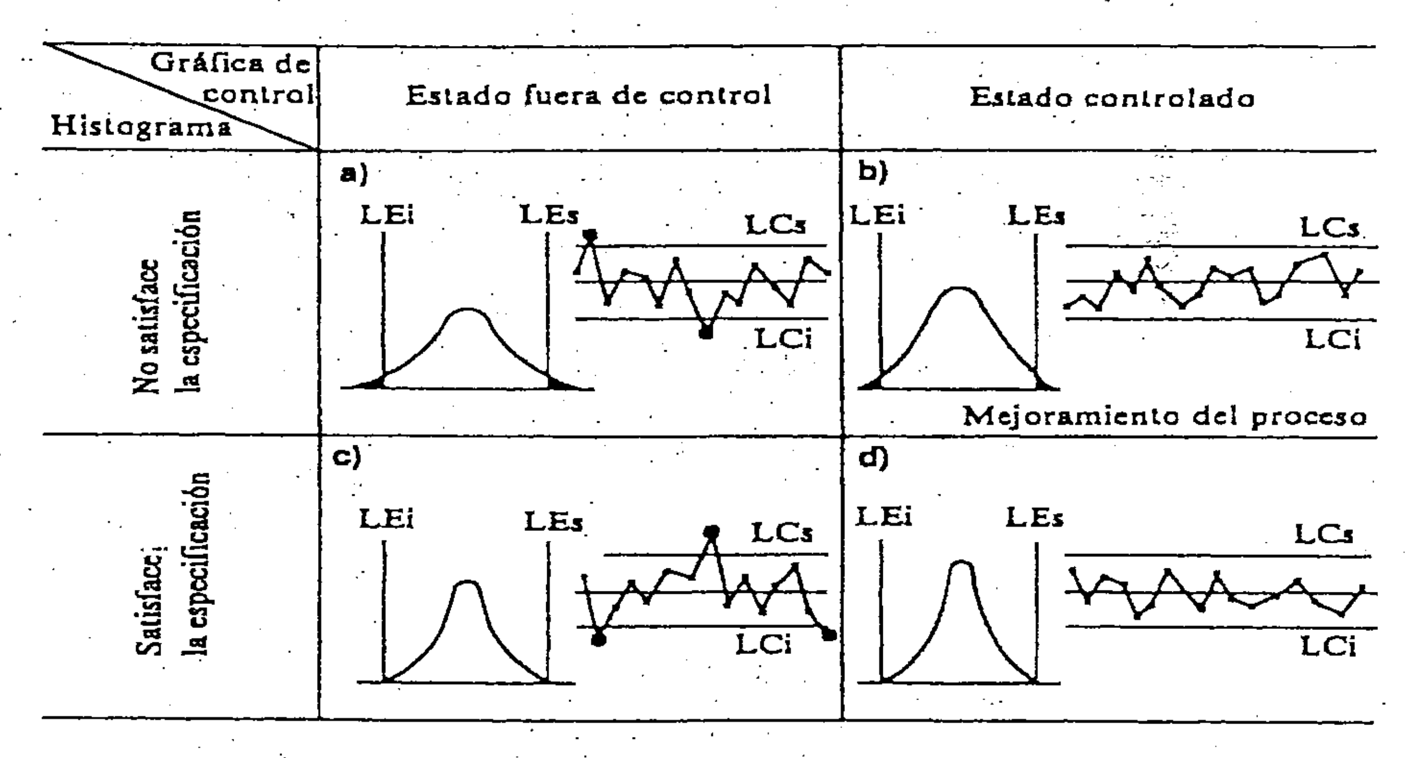


Figura 7.20 Comparación de gráficas de control con las especificaciones

No hay problema en el caso d), pero se presenta confusión cuando ocurren b) o c). La razón es una falta de armonía entre la capacidad del proceso y la especificación. En el caso b), la capacidad del proceso es insuficiente para la especificación. Para prevenir la ocurrencia de productos defectuosos, debe hacerse el esfuerzo para mejorar la capacidad del proceso. El caso c) ocurre cuando el proceso tiene suficiente capacidad para la especificación. En este caso, debe considerarse el mejoramiento de la eficiencia de la producción.

Ejercicio 7.1

Seleccione un tipo apropiado de gráfica de control para controlar las siguientes características de control.

1) Peso de galletas empacadas.
2) El número de productos defectuosos en 1.000 partes.
3) El número de defectos de soldadura en un equipo de radio.
4) Rendimiento de un producto químico en una serie.
5) El porcentaje de productos defectuosos en un lote cuyo tamaño puede variar.
6) El valor de resistencia de cinco piezas de prueba muestreadas en un día.
7) El número de rayones por 1 m^2 en una placa de acero.

Ejercicio 7.2

Para investigar el perfil de variación de ciertas partes en un proceso de maquinado, se midieron las dimensiones de las partes cuatro veces al día, a las 9:00, 11:00 y 16:00. Haga una gráfica de control $\bar{x} - R$ para analizar este proceso.



No.	Fecha	Hora			
		9:00	11:00	14:00	16:00
1	Nov. 2	52.5	52.9	52.9	53.5
2	3	53.0	52.8	53.5	52.4
3	4	52.8	52.9	52.7	52.8
4	5	52.9	52.9	52.9	52.9
5	6	52.8	52.9	52.7	53.1
6	9	52.6	53.4	53.1	53.3
7	10	53.5	53.6	52.8	52.7
8	11	53.1	53.3	53.5	53.0
9	12	53.4	53.1	53.1	53.1
10	13	53.2	53.4	53.1	52.9
11	16	53.4	53.0	53.9	53.1
12	17	52.8	52.9	53.2	53.2
13	18	53.2	53.3	52.9	53.1
14	19	53.5	52.9	54.0	53.9
15	20	54.3	53.6	53.6	53.8
16	23	53.2	53.3	54.0	53.7
17	24	53.8	54.0	53.8	53.8
18	25	53.1	53.6	53.7	53.8
19	26	53.7	53.8	53.0	53.5
20	27	53.3	53.1	53.6	53.0
21	30	53.3	53.7	53.3	53.8
22	Dic. 1	53.1	53.1	53.2	53.1
23	2	53.6	53.4	53.2	53.0
24	3	53.4	53.7	53.0	53.2
25	4	53.3	53.2	53.5	53.4

Ejercicio 7.3

Corrija los errores, cuando sea necesario, en las siguientes ocasiones.

1) En una gráfica $\bar{x} - R$, la gráfica $\bar{x}$ muestra un cambio en la media de un subgrupo y la gráfica R muestra un cambio en la variación en el interior de un subgrupo.
2) En una gráfica $\bar{x} - R$, miramos si hay variación entre subgrupos con base en la variación dentro de los subgrupos.
3) Cuando todos los puntos registrados en una gráfica de control se encuentran entre los límites de control, consideramos que el proceso está en un estado de control.
4) En una gráfica $\bar{x} - R$, un punto caerá por fuera de los límites, aun si no hay cambio en el proceso.
5) En una gráfica de control en la cual es deseable un valor pequeño de la característica de control, en la práctica no se utiliza el LCi.
6) Cuando una gráfica de control muestra un estado controlado, el proceso producirá productos que se conforman con la especificación.
7) Suponga que un punto cae por fuera de los límites de control. Si los límites de especificación son amplios, sólo tenemos que cuidarnos del proceso. Pero si los límites de la especificación son estrechos, debemos averiguar la causa prontamente y tomar acciones apropiadas.
8) Cuando una gráfica $\bar{x} - R$ muestra un estado estadísticamente controlado, podremos captar la capacidad del proceso por medio de un histograma de x (datos básicos).
9) En una gráfica $\bar{x}$ hay quince puntos sucesivos entre los límites de control y $\pm$ 1-sigma, el proceso está en un estado de perfecta estabilidad. Debemos tratar de mantener este estado.



VIII
Aditividad de las varianzas

8.1 LAS MEDIAS Y LAS VARIANZAS DE SUMAS

Antes de tirar un dado no sabemos qué cara quedará hacia arriba, pero si el dado no está cargado, la probabilidad de que una cualquiera de las caras quede hacia arriba es de 1/6. El valor esperado del conjunto de puntos es

$$\mu = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5 \ .$$

La varianza del número de puntos es

$$\sigma^2 = \sum_{x=1}^{6} (x - 3.5)^2 \times \frac{1}{6} = \frac{35}{12} \ , \qquad (8.1)$$

y su desviación estándar es

$$\sigma = \sqrt{\frac{35}{12}} = 1.71 \ . \qquad (8.2)$$

Cuando se tira un dado dos veces y se obtiene la suma del número de puntos, será algún valor entre 2 y 12, y sus frecuencias relativas pueden obtenerse de la tabla 8.1. En este caso, los



valores aparecen con frecuencias diferentes, siendo 7 el más frecuente y 2 y 12 los menos frecuentes.

¿Cuáles son la media y la varianza de las sumas de los números de puntos en este caso?

Primero, la media es

$$\mu = \sum xP(x)$$
$$= 2\times\frac{1}{36} + 3\times\frac{2}{36} + \cdots + 6\times\frac{5}{36} + 7\times\frac{6}{36}$$
$$+ 8\times\frac{5}{36} + \cdots + 12\times\frac{1}{36} = 7\ . \qquad (8.3)$$

La varianza es

$$\sigma^2 = \sum(x-\mu)^2 P(x)$$
$$= (2-7)^2\times\frac{1}{36} + (3-7)^2\times\frac{2}{36} + \cdots$$
$$+ (7-7)^2\times\frac{6}{36} + \cdots + (12-7)^2\times\frac{1}{36} = \frac{35}{6}\ .$$
$$(8.4)$$

Suma del número de puntos		Segunta tirada					
		1	2	3	4	5	6
Primera tirada	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Tabla 8.1 Suma del número de puntos obtenidos en dos tiros del dado

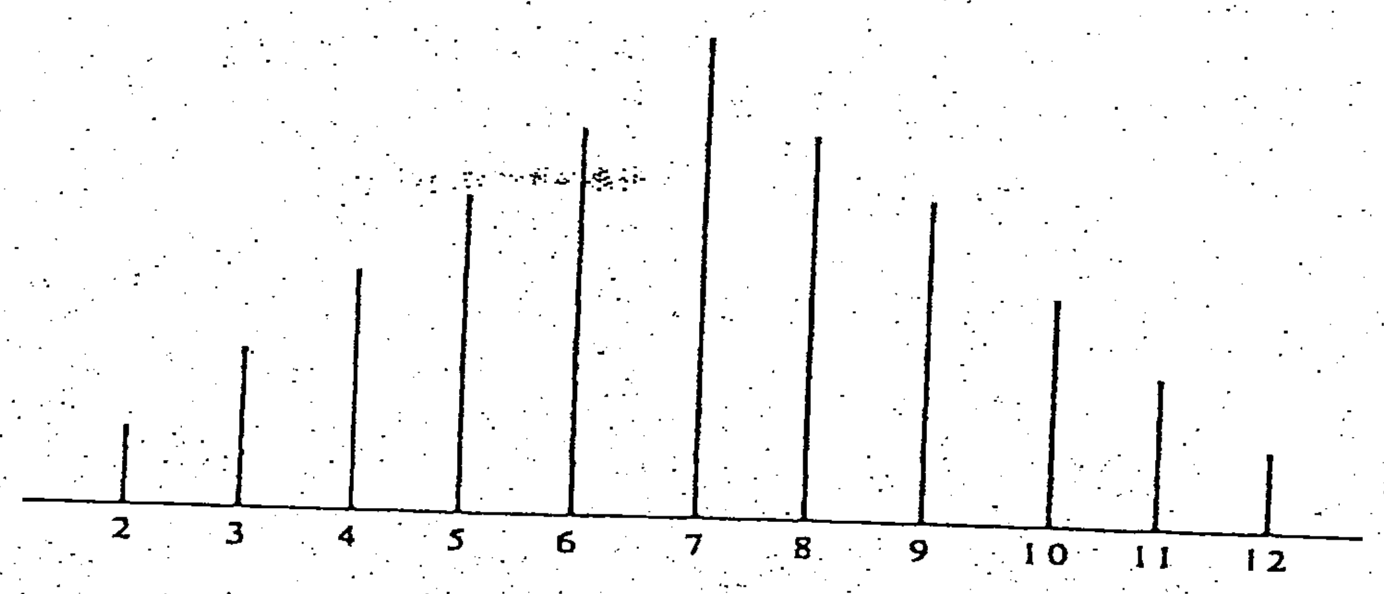


Figura 8.1 Frecuencias relativas de la suma del número de puntos obtenidos cuando un dado se tira dos veces

De esto podemos ver que la media y la varianza de las sumas de los números de puntos obtenidos en dos tiradas de un dado son dos veces los valores obtenidos en una tirada.

Después, hagamos lo mismo para la diferencia en el número de puntos. Es decir, ¿cuál será el valor de la diferencia cuando el segundo valor se resta del primero? Como podemos ver en la tabla 8.2, los datos consisten en valores entre –5 y +5, y en lo referente a las frecuencias relativas, 0 es la más frecuente, y las frecuencias relativas de cada valor son como se indica en la figura 8.2, que es la misma figura 8.1 excepto que los números se han corrido siete puestos a la izquierda. A partir de esto sabemos que la media y la varianza de las diferencias son

$$\mu = 0\,, \tag{8.5}$$

y

$$\sigma^2 = \frac{35}{6}\,. \tag{8.6}$$

Nótese que las varianzas de las sumas y las diferencias son las mismas.

Diferencia entre la primera y segunda tirada		Segunda tirada					
		1	2	3	4	5	6
Primera tirada	1	0	−1	−2	−3	−4	−5
	2	1	0	−1	−2	−3	−4
	3	2	1	0	−1	−2	−3
	4	3	2	1	0	−1	−2
	5	4	3	2	1	0	−1
	6	5	4	3	2	1	0

Tabla 8.2 Diferencia entre la primera tirada y la segunda tirada

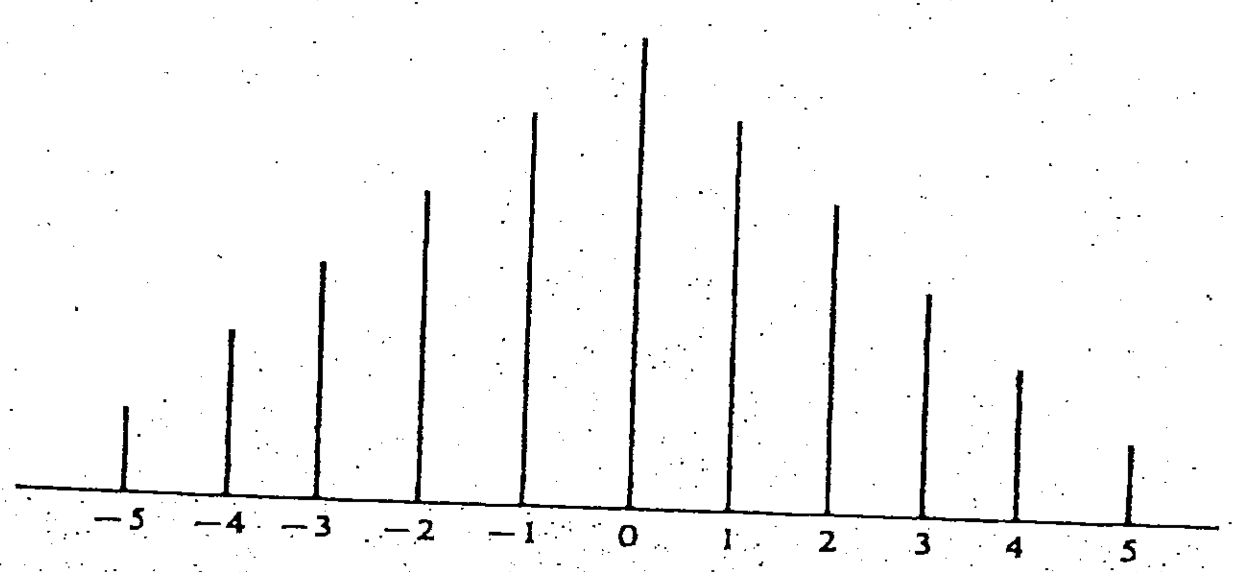


Figura 8.2 Frecuencias relativas de la diferencia entre el número de puntos obtenidos después de la primera y la segunda tirada del dado

8.2 PRECISIÓN DEL ENSAMBLAJE DE PARTES

Supongamos que tenemos un producto que, como lo muestra la figura 8.3a, se hace uniendo las partes *A* y *B*. Cuando la dimensión x de la parte *A* tiene una distribución de media μ_x, y una desviación estándar de σ_x y la dimensión y de la parte *B* tiene una

distribución de la media μ_y y una desviación estándar σ_y, si se toman muestras aleatorias de las dos partes y se unen, entonces la media y la varianza de la dimensión z serán:

$$\mu_z = \mu_x + \mu_y \,, \tag{8.7}$$

y

$$\sigma_z^{\,2} = \sigma_x^{\,2} + \sigma_y^{\,2}. \tag{8.8}$$

Cuando se toman aleatoriamente dos partes y se las une como en la figura 8.3b, tenemos

$$\mu_z = \mu_x - \mu_y \,, \tag{8.9}$$

y

$$\sigma_z^{\,2} = \sigma_x^{\,2} + \sigma_y^{\,2} \tag{8.10}$$

para la dimensión z. La media de z está dada por la diferencia entre las medias de A y de B. Pero la varianza de z es la suma de las varianzas de las dos.

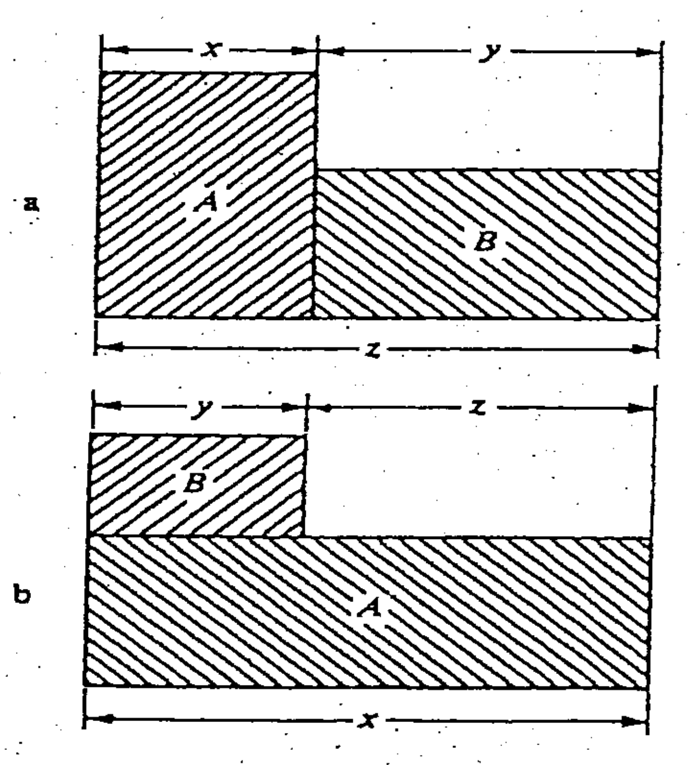


Figura 8.3 Ensamblaje de partes

8.3 FÓRMULAS TEÓRICAS

Cuando z se define por

$$z = ax + by \tag{8.11}$$

cuando a y b son coeficientes constantes, la media de z está dada por

$$\mu_z = a\mu_x + b\mu_y \,, \tag{8.12}$$

donde μ_x y μ_y son la media de x y de y. Si x y y son independientes, la varianza de z está dada por

$$\sigma_z{}^2 = a^2\sigma_x{}^2 + b^2\sigma_y{}^2 \,, \tag{8.13}$$

donde σ_x^2 y σ_y^2 son las varianzas de x y de y.

En las partes ensambladas descritas en la sección previa, $a = 1$ y $b \pm 1$. Debido a que se suman los cuadrados de los coeficientes, la varianza se convierte en una suma aun cuando el caso particular implica una diferencia de variables aleatorias. Esta característica se llama la *aditividad de las varianzas.*

Generalmente, si $x_1, x_2, \ldots, x_n$, son variables aleatorias independientes y sus medias y varianzas son $\mu_1, \ldots, \mu_n$ y $\sigma_1^2, \ldots, \sigma_n^2$, respectivamente, la expectativa y la varianza de

$$y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \tag{8.14}$$

están dadas por

$$\mu_y = a_1 \mu_1 + \ldots + a_n \mu_n \,, \tag{8.15}$$

y

$$\sigma_y^2 = a_1{}^2\sigma_1{}^2 + \cdots + a_n{}^2\sigma_n{}^2 \,. \tag{8.16}$$



8.4 EL VALOR ESPERADO Y LA VARIANZA DE LA MEDIA MUESTRAL

Cuando se toman n mediciones de una población con una media poblacional μ y una varianza poblacional σ^2, donde los valores de las mediciones son $x_1, x_2, \ldots, x_n$ y su media es y, entonces

$$y = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n \,. \tag{8.17}$$

De la fórmula anterior, poniendo $a_1 = a_2 = \ldots a_n = 1/n$, la expectativa y la varianza de y se obtienen por

$$\mu_y = \mu \,, \tag{8.18}$$

y

$$\sigma_y^2 = (\frac{1}{n})^2 n\, \sigma^2 = \frac{\sigma^2}{n} \,. \tag{8.19}$$

Ésta es una fórmula bien conocida para la distribución de la media muestral.

8.5 ERROR DE MUESTREO Y ERROR DE MEDICIÓN

Cuando se estima un ingrediente de materiales tales como el carbón o el mineral de hierro, el trabajo se hace generalmente en dos etapas, la primera de las cuales es tomar una muestra del lote, y la segunda es el análisis químico de la muestra obtenida. Debido a estos procedimientos, existen dos tipos de errores en la estimación del ingrediente. Uno es llamado el error de muestreo y ocurre cuando la muestra se toma del lote, y el otro se llama error de medición, el cual ocurre cuando se mide la muestra. Cuando el error de muestreo se expresa como x_S y el error de medición como x_M, los datos medidos y se pueden expresar como

$$y = \mu + x_S + x_M \,, \tag{8.20}$$

donde μ es el verdadero valor del lote. Si la media de x_S es μ_S y la de x_M es μ_M, la media de y está dada por

$$\mu_y = \mu + \mu_S + \mu_M, \tag{8.21}$$

donde μ_S μ_M son el sesgo en el muestreo y la medición respectivamente. La varianza de y está dada por

$$\sigma_y^{\ 2} = \sigma_S^{\ 2} + \sigma_M^{\ 2}, \tag{8.22}$$

donde σ_s^2 es la precisión de muestreo y σ_M^2 es la precisión de medida.

Cuando una muestra se mide dos veces y se toma la media, la varianza de la media es

$$\sigma_{\bar{y}}^{\ 2} = \sigma_S^{\ 2} + \frac{\sigma_M^{\ 2}}{2}. \tag{8.23}$$

La varianza del error de medición se convierte en la mitad. Si se toman dos muestras, haga mediciones de cada una de ellas una vez y tome la media, entonces la varianza está dada por

$$\sigma_{\bar{y}}^{\ 2} = \frac{\sigma_S^{\ 2}}{2} + \frac{\sigma_M^{\ 2}}{2}, \tag{8.24}$$

y en este caso la varianza del error de muestreo también es la mitad.

Como se explicó en la sección previa, la varianza del valor de la media se convierte en $1/n$, si las mediciones se repiten n veces, pero si ocurren errores en diversas etapas como en este ejemplo, los componentes de la varianza se reducen de diversas maneras dependiendo del estado en el cual se hace la repetición.

Los errores pueden clasificarse en dos categorías, sesgo y variación. Mientras que la variación se reduce cuando se repiten las mediciones, el sesgo permanece igual sin importar cuántas observaciones se repiten, y la obtención de la media no lleva a su reducción.

8.6 LA VARIANZA DE LOS VALORES DE UNA FUNCIÓN

Cuando y se expresa por la función f de variables aleatorias independientes $x_1, x_2, \ldots, x_n$, como

$$y = f(x_1, x_2, \cdots, x_n), \tag{8.25}$$

y si f se amplía alrededor de $\mu_1, \cdots, \mu_n$) que son los valores medios de $x_1, \cdots, x_n$, entonces

$$\begin{aligned} y &= f(x_1, \cdots, x_n) \\ &= f(\mu_1, \cdots, \mu_n) + \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)(x_i - \mu_i) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_i - \mu_i)(x_j - \mu_j) + \cdots \end{aligned} \tag{8.26}$$

Si los términos de la función pueden despreciarse a partir de la segunda derivada, es posible escribir

$$\begin{aligned} y &\doteqdot f(\mu_1, \cdots, \mu_n) \\ &\quad + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - \mu_i) \end{aligned} \tag{8.27}$$

y en consecuencia la media y la varianza de y son aproximadamente

$$\mu_y \doteqdot f(\mu_1, \cdots, \mu_n) \tag{8.28}$$

y

$$\sigma_y{}^2 \doteqdot \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i{}^2, \tag{8.29}$$

donde σ_i^2 es la varianza de x_i.

Ejemplo 8.1

Peso del carbón seco

Se entregaron alrededor de 1.000 toneladas de carbón. Su contenido promedio de humedad es alrededor de 8%. Si la precisión en el pesaje del carbón, σ_1, es 5 toneladas y la precisión del muestreo y de la medición de la humedad, σ_2, es 0.5%, ¿cuál es la precisión, σ_y, de y, el peso estimado del carbón cuando está seco?

Sea el peso medido del carbón x_1 toneladas y el contenido de humedad sea x_2%, entonces el peso del carbón seco se estima por

$$y = x_1 (1 - \frac{x_2}{100}) . \tag{8.30}$$

En consecuencia,

$$\begin{aligned} \sigma_y{}^2 &\fallingdotseq (1 - \frac{x_2}{100})^2 \sigma_1{}^2 + x_1{}^2 (\frac{\sigma_2}{100})^2 \\ &= (\frac{92}{100})^2 \times 5^2 + 1000^2 (\frac{0.5}{100})^2 = 46.16 . \end{aligned} \tag{8.31}$$

Así

$$\sigma_y = 6.8 \ \text{(toneladas)} \tag{8.32}$$

8.7 CUANDO LAS VARIABLES ALEATORIAS NO SON INDEPENDIENTES

La aditividad de las varianzas funciona bien cuando las variables aleatorias son mutuamente independientes. Se dice que dos variables aleatorias son independientes cuando el valor de una variable cambia sin ninguna relación con la otra variable.

Cuando hay n parejas de casados y las edades de un esposo y su esposa son x y y, respectivamente, la diferencia de sus edades se expresa por

$$z = x - y . \tag{8.33}$$

Si x y y son independientes, la varianza z será

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2. \tag{8.34}$$

Suponga $\sigma_x = \sigma_y = 5$, entonces $\sigma_z = 7$. Pero en realidad, esto no sucede, porque las edades de un esposo y de su esposa no son independientes. En una pareja anciana, tanto el esposo como la esposa son ancianos, y en una pareja joven ambos son jóvenes. Esto no quiere decir que no haya hombres ancianos casados con mujeres jóvenes ni mujeres ancianas casadas con hombres jóvenes, pero estos casos son mucho menos frecuentes que los dos primeros casos.

Cuando x y y no son independientes, la media y la varianza de

$$z = ax + by \tag{8.35}$$

están dadas por

$$\mu_z = a\mu_x + b\mu_y , \tag{8.36}$$

y

$$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y . \tag{8.37}$$

La media de z es la misma como en el caso en el que x y y son independientes, pero la varianza difiere por $2ab\rho\sigma_x\sigma_y$, donde ρ es el coeficiente de correlación, coeficiente que muestra el grado de relación entre dos variables. Los valores de ρ se encuentran entre -1 y $+1$, es decir

$$-1 \leq \rho \leq 1 . \tag{8.38}$$

A medida que la relación entre las dos variables es más fuerte, más cercano es el valor absoluto de ρ a 1.

Cuando $\rho = \pm 1$, todos los valores (x, y) se encuentran en una línea recta. Cuando $\rho > 0$, el aumento de una tiende a aumentar el valor de la otra y se llama una correlación positiva, y cuando $\rho < 0$, el aumento de una tiende a disminuir el valor de la otra y se llama una correlación negativa. En el caso de las edades de los hombres y sus esposas, claramente $\rho > 0$. Por ejemplo, si $\rho = 0.7$ la varianza de las diferencias de edad serán

$$\sigma_z{}^2 = 5^2 + 5^2 - 2 \times 0.7 \times 5 \times 5 = 15, \quad (8.39)$$

$$\sigma_z \doteqdot 3.87. \quad (8.40)$$

8.8 COMBINACIÓN SELECTIVA

Cuando las partes se seleccionan aleatoriamente y se ensamblan, las varianzas de las sumas y las diferencias de las dimensiones de estas partes son cada una la suma de las varianzas de cada parte debido a la aditividad de las varianzas, como se describió en la sección 8.2, y la variación será siempre mayor que la de las partes originales. La manera ortodoxa de reducir la variación de un producto es reducir las variaciones de sus partes individuales. Sin embargo, si por razones tecnológicas de la producción es demasiado difícil hacerlo, es necesario seleccionar la "pareja" más adecuada para una parte, en lugar de hacerlo aleatoriamente.

Por ejemplo, la distancia entre el pistón de un motor y su cilindro tiene un gran efecto en el comportamiento del motor. Si el diámetro del cilindro es x y el diámetro del pistón es y, es necesario mantener el valor $z = x - y$ tan uniforme como sea posible. A partir de la fórmula descrita en la sección 8.7, obtenemos

$$\sigma_z{}^2 = \sigma_x{}^2 + \sigma_y{}^2 + 2\rho\sigma_x\sigma_y. \quad (8.41)$$

De manera que si ρ se hace aproximar a 1, $\sigma_z{}^2$ se aproximará a $(\sigma_x - \sigma_y)^2$.



La lente de una cámara se hace laminando lentes de diferentes índices de refracción con el fin de reducir la aberración cromática. Si el espesor de una lente es x y el de la otra lente es y, es deseable mantener el espesor general $z = x + y$ tan uniforme como sea posible.

Puesto que

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y,$$

σ_z^2 se acercará a $(\sigma_x - \sigma_y)^2$, si se hace que ρ se aproxime a -1, como en el ejemplo anterior.

Aunque la combinación selectiva es una manera afectiva de reducir la variación, se disminuye la eficiencia de la producción porque se requieren más partes y cada parte tiene que medirse por anticipado.

8.9 CONTROL ESTADÍSTICO DE CALIDAD

La producción industrial implica frecuentemente la producción masiva del mismo tipo de producto. Es necesario mantener la variación de las características de calidad de estos productos en el mínimo, y lograrlo es una de las tareas más importantes del control de calidad. Las variaciones de las características de calidad son causadas por cambios en un gran número de factores que afectan a estas características. Éstos pueden generalmente clasificarse en los siguientes cuatro elementos.

1) Variaciones en los materiales.
2) Variaciones en la maquinaria y los equipos.
3) Variaciones humanas y en métodos (trabajadores y métodos de operación).
4) Variaciones en las medidas.

Éstos se llaman los 4M's de las variaciones, y las variaciones en las características de calidad se presentan como la suma de estos cuatro tipos de variaciones según el principio de la aditividad de



las varianzas. Debemos averiguar qué tan fuertemente contribuyen estos elementos a las variaciones en la calidad y cuáles de estas variaciones se deben controlar, y diseñar métodos para controlarlas. Las actividades básicas del control de calidad en la fábrica son los análisis repetidos y las mejorías para reducir la variación de la calidad. Empezando con esto, es necesario determinar la magnitud de las actuales fluctuaciones, y luego proceder con los análisis de los factores que las causan.

A partir del teorema de la aditividad de las varianzas, podemos derivar una importante ley, de la manera siguiente:

"Reduzca la varianza mayor con el fin de reducir la varianza total".

La tabla 8.3 muestra varios casos de valores de

$$\sigma_x,\ y\ \sigma_y,\ y\ \sigma_z^{\,2} = \sigma_x^{\,2} + \sigma_y^{\,2}.$$

Caso No.	σ_x	σ_y	$\sigma_z^{\,2} = \sigma_x^{\,2} + \sigma_y^{\,2}$	σ_z
1	10	10	200	14.1
2	10	5	125	11.2
3	10	3	109	10.4
4	10	1	101	10.05
5	5	5	50	7.07
6	5	3	34	5.83
7	5	1	26	5.10

Tabla 8.3 La reducción de la mayor varianza para reducir la varianza total

Podemos ver que la contribución de σ_x a σ_z es mucho mayor que la de σ_y, excepto cuando $\sigma_x = \sigma_y$. Esto significa que debemos primero reducir la mayor varianza $\sigma_x^{\,2}$, cuando queremos reducir la varianza total $\sigma_z^{\,2}$.



Ejercicio 8.1

Se seleccionan aleatoriamente y se ensamblan dos tipos de partes, A y B. Los siguientes datos muestran el valor medido de cada parte antes del ensamblaje.

1) Obtenga las varianzas de A y de B.
2) Obtenga el valor de $A + B$ para cada conjunto y calcule la varianza.
3) Haga el mismo cálculo para $A - B$.

No.	A	B	No.	A	B
1	6.95	5.40	6	7.70	3.90
2	6.75	4.45	7	6.85	4.25
3	7.25	4.65	8	7.50	3.95
4	6.50	4.55	9	7.05	4.80
5	7.95	4.95	10	7.90	4.90

Ejercicio 8.2

Una máquina empacadora detiene automáticamente su operación de empaque cuando la suma del peso del contenedor y su contenido alcanza un valor prefijado. ¿Cuál es la variación del peso del contenido cuando la precisión de empaque de esta máquina en términos de la desviación estándar es σ_1 y la variación del peso del contenedor es σ_2 en términos de la desviación estándar?

IX

Introducción a la inferencia estadística

9.1 ESTADÍSTICA

Cuando queremos estimar la media de una población, tomamos un conjunto de observaciones de esa población y calculamos la media de las observaciones. Un valor calculado a partir de una muestra, tal como la media de la muestra, se llama una *estadística*. En otras palabra, una estadística es una función de observaciones de la muestra.

Debemos diferenciar entre una estadística y un parámetro de población. Para hacerlo, usamos con frecuencia los términos *media de la muestra* y *media de la población* en lugar de usar simplemente media. Un parámetro de la población tiene un cierto valor constante, pero no se conoce en realidad. Por otra parte, podemos calcular una estadística a partir de una muestra, pero la estadística variará de muestra a muestra. Aunque quisiéramos conocer el parámetro de la población, observamos solamente muestras obtenidas de la población. Por tanto, tenemos que estimar el parámetro de la población a partir de una estadística. Para hacerlo, tenemos que conocer la distribución de las estadísticas, tales

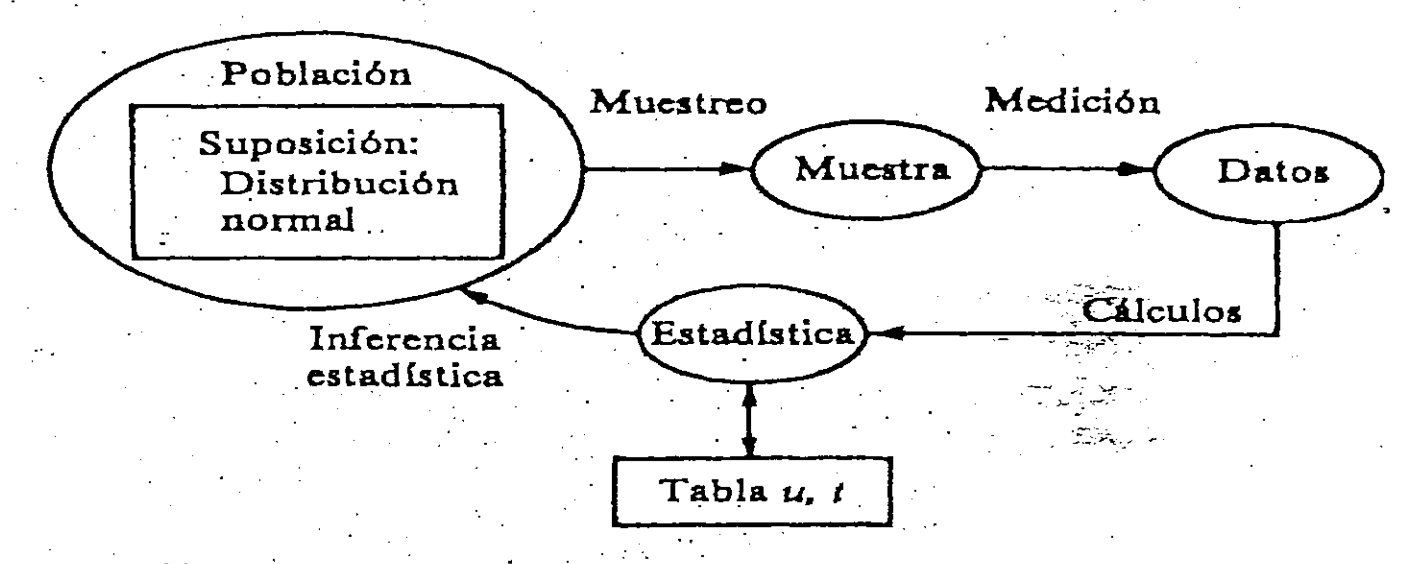


Figura 9.1 Inferencia estadística

como la distribución u o la distribución t. La figura 9.1 ilustra el método de la inferencia estadística.

9.2 DISTRIBUCIÓN DE LAS ESTADÍSTICAS

En la inferencia estadística, debemos conocer la distribución de varias estadísticas. Éstas se pueden derivar de un experimento de muestreo. Usando un computador, generamos un conjunto de 5 variables aleatorias desde $N(50, 2^2)$, y calculamos $\bar{x}$, R, V, s, u y t donde

$$u = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{x} - 50}{2 / \sqrt{5}} \tag{9.1}$$

y

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{\bar{x} - 50}{s / \sqrt{5}}. \tag{9.2}$$

Repetimos los anteriores cálculos 10 000 veces. De esta manera tendremos 10 000 observaciones de las estadísticas $\bar{x}$, R y

$\bar{x}_1, \cdots, x_5 \sim N(50, 2^2)$

No.	X_1	X_2	X_3	X_4	X_5	$\bar{X}$	R	V	s	u	t
1	51	55	50	47	50	50.6	8	8.30	2.88	0.67	0.47
2	52	50	49	49	51	50.2	3	1.70	1.30	0.22	0.34
3	47	51	49	52	46	49.0	6	6.50	2.55	−1.12	−0.88
4	51	50	50	47	51	49.8	4	2.70	1.64	−0.22	−0.27
5	50	51	47	51	47	49.2	4	4.20	2.05	−0.89	−0.87
6	50	53	49	51	54	51.4	5	4.30	2.07	1.57	1.51
7	49	50	50	51	51	50.2	2	0.70	0.84	0.22	0.53
8	46	47	47	49	50	47.8	4	2.70	1.64	−2.46	−2.99
9	48	49	50	50	50	49.4	2	0.80	0.89	−0.67	−1.50
10	52	50	49	53	51	51.0	4	2.50	1.58	1.12	1.41
11	49	51	51	48	53						
12	52	51	54	50	52						
13	50	49	48	52	51						
14	51	49	52	50	48						
15	51	51	49	53	48						
16	51	55	47	50	51						
17	53	48	49	53	50						
18	53	50	49	50	51						
19	46	50	50	52	50						
20	53	52	50	50	48						

Tabla 9.1 Parte de los resultados de un experimento de muestreo



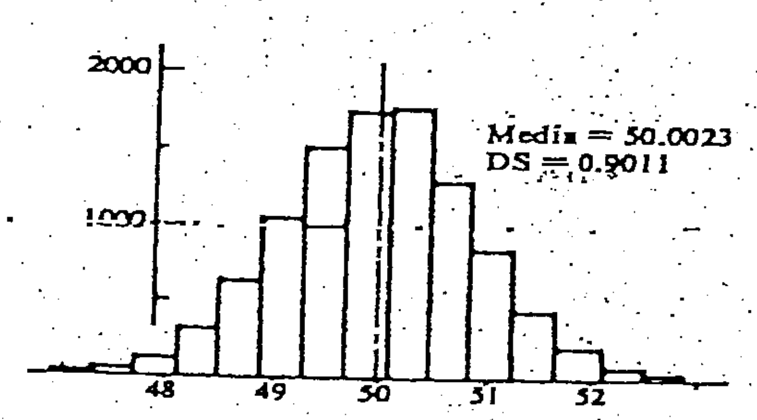


Figura 9.2 Distribución de $\overline{x}$

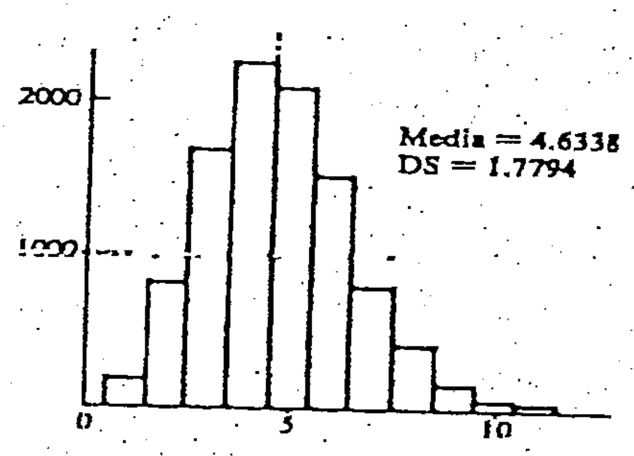


Figura 9.3 Distribución de R

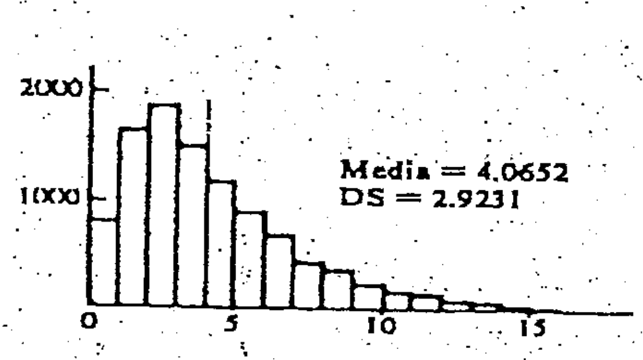


Figura 9.4 Distribución de V

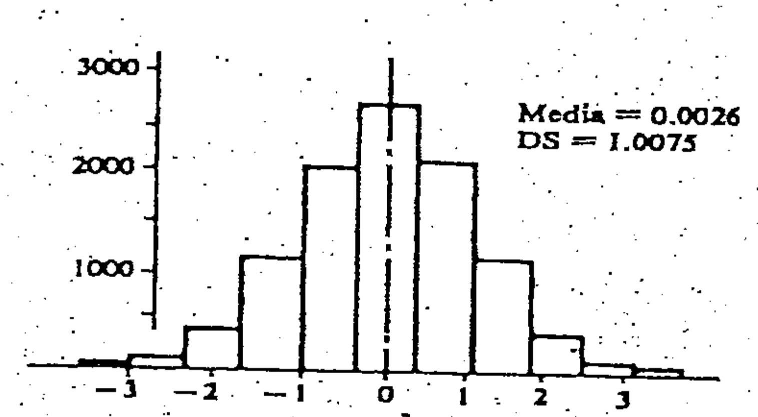


Figura 9.5 Distribución u

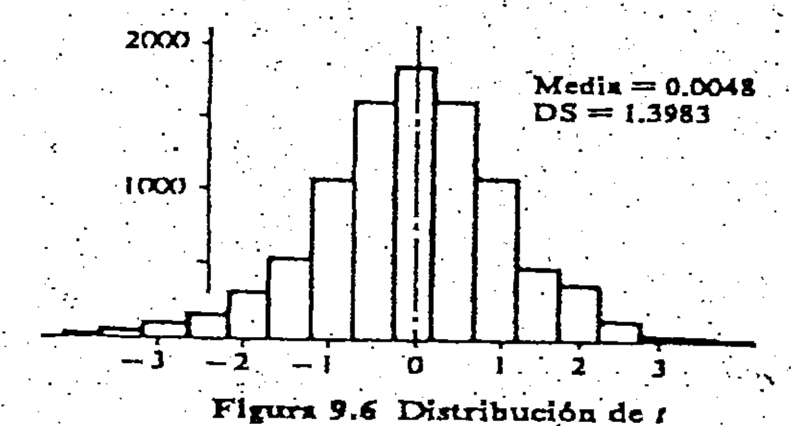


Figura 9.6 Distribución de t

así sucesivamente. La tabla 9.1 muestra los primeros 20 casos. Los espacios en blanco en los números 11-20 se dejan como ejercicio para el lector. Los histogramas de las 10 000 observaciones de las estadísticas se muestran en las figuras 9.2 a 9.6, los cuales muestran el contorno de la distribución.

(1) Distribución de $\bar{x}$

Por la figura 9.2 podemos ver que:

a) x se distribuye simétricamente alrededor de la media, dando la apariencia de una distribución normal.
b) La media de $\bar{x}$ (= 50.0023) se acerca mucho a la media de la población (= 50.0).
c) La desviación estándar de $\bar{x}$ (= 0.9011) se acerca a $1/\sqrt{n}$ veces la desviación estándar de la población $(2.0/\sqrt{5} = 0.8944)$.

Por lo general, el siguiente teorema se cumple:

Teorema 9.1

Supongamos que $x_1, \ldots, x_n$ son n observaciones de una población con una media μ y una varianza σ^2, y que $\bar{x}$ es la media de la muestra. Entonces, el valor esperado, la varianza y la desviación estándar de $\bar{x}$ son

$$E(\bar{x}) = \mu , \tag{9.3}$$

$$V(\bar{x}) = \frac{\sigma^2}{n} , \tag{9.4}$$

y

$$D(\bar{x}) = \frac{\sigma}{\sqrt{n}} , \tag{9.5}$$

respectivamente.



Teorema 9.2

Supongamos que $(x_1, \ldots, x_n)$ de $N(\mu, \sigma^2)$, y $\bar{x}$ es la media de la muestra. Entonces, $\bar{x}$ se distribuye como $N(\mu, \sigma^2/n)$.

El teorema 9.1 se deriva directamente de la sección 8.4. Es un hecho importante que la desviación estándar de una media de la muestra sea $1/\sqrt{n}$ veces la de la población (figura 9.7). La precisión de la media de la muestra $\bar{x}$ como estimador de la media de la población es proporcional a la raíz cuadrada del tamaño de la muestra, $\sqrt{n}$.

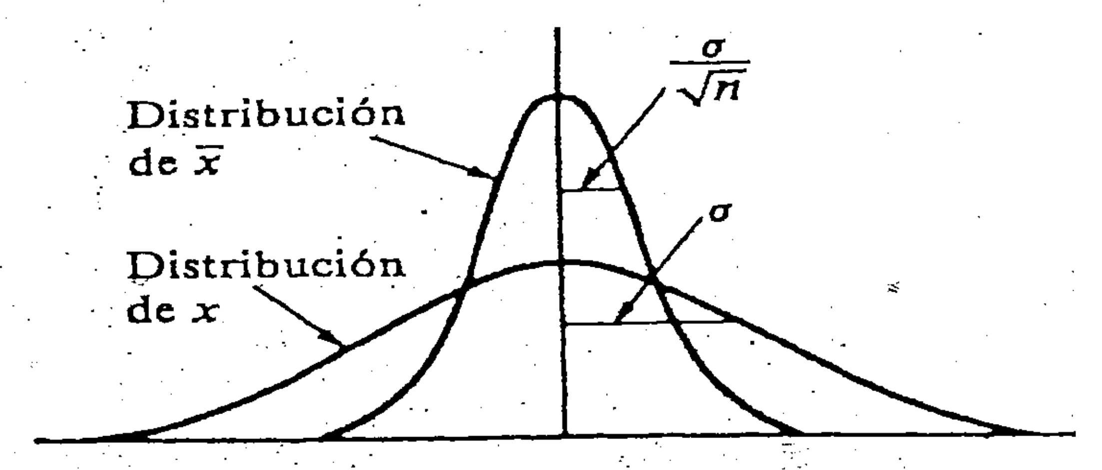


Figura 9.7 Distribución de $\bar{x}$ y de x

Otro teorema importante sobre la media de la muestra es el *teorema del límite central*, enunciado de la manera siguiente:

Teorema 9.3 (Teorema del límite central)

La distribución de una media muestral de una población que tiene una varianza finita tiende a distribuirse normalmente a medida que el tamaño de la muestra tiende hacia el infinito.



Figura 9.8 Ejemplos del teorema del límite central

Según el teorema 9.2, la media muestral de una población distribuida normalmente se distribuye exactamente en forma de una distribución normal. Y el teorema 9.3 dice que, aun si la distribución de una población no es normal, la media muestral se aproxima a la distribución normal. Esta aproximación se cumple mejor para n de valor grande, pero es adecuada para un valor de n tan bajo como 5. Por ejemplo, la figura 9.8 muestra la distribución de una media muestral de n observaciones tomadas de

una distribución uniforme. Se debe al teorema de límite central que podamos hacer varias inferencias estadísticas usando la media de la muestra, suponiendo que la población está distribuida normalmente.

(2) Distribución de R

En la figura 9.3 podemos ver que:

a) La distribución tiene un sesgo positivo.
b) La media (= 4.6338) es alrededor de 2.3 veces la desviación estándar de la población.
c) La desviación estándar (= 1.7794) es alrededor de 0.9 veces la desviación estándar de la población.

Por lo general, el siguiente teorema se cumple:

Teorema 9.4

Sea R el rango de una muestra $(x_1, \ldots, x_n)$ de $N(\mu, \sigma^2)$. El valor esperado y la desviación estándar de R son

$$E(R) = d_2\sigma, \tag{9.6}$$

y

$$D(R) = d_3\sigma, \tag{9.7}$$

respectivamente, donde d_2 y d_3 son ciertas constantes que dependen de n.

Los valores de d_2 y d_3 se muestran en la tabla A.2 del apéndice de este libro. A partir de la tabla, en este caso, $n = 5$, $d_2 = 2.326$ y $d_3 = 0.864$. A partir del teorema podemos estimar σ por

$$\hat{\sigma} = \frac{R}{d^2}. \tag{9.8}$$

(3) Distribución de V

Por la figura 9.4 vemos que:

a) La distribución tiene un sesgo positivo.
b) La media (= 4.0652) es casi igual a la varianza de la población (2.0^2 = 4.0).
c) La desviación estándar (= 2.9231) es aproximadamente 0.7 veces la varianza de la población.

Por lo general, el siguiente teorema se cumple:

Teorema 9.5

Sea V la varianza de una muestra $(x_1, \ldots, x_n)$ de $N(\mu, \sigma^2)$. El valor esperado y la desviación estándar de V son:

$$E(V) = \sigma^2, \tag{9.9}$$

y

$$D(V) = \sqrt{\frac{2}{n-1}}\,\sigma^2, \tag{9.10}$$

Se usa la estadística V para estimar la varianza de la población σ^2. A partir de (9.9), su valor esperado es igual a σ^2. Un estimador cuyo valor esperado es igual al parámetro estimado de la población se llama un *estimador insesgado*. Así, V es un estimador no sesgado de σ^2. Por supuesto que $\bar{x}$ también es un estimador no sesgado de μ.

A propósito, V se obtiene dividiendo la suma de los cuadrados de las desviaciones S por $(n - 1)$, como se define en (5.4). Puede parecer extraño dividir S por $(n - 1)$ en lugar de dividir por n, y una razón para hacerlo es la propiedad expresada en (9.9). La otra razón es que el número de variables independientes usadas en el cálculo de S es $(n - 1)$. Por ejemplo, cuando

$n = 1$, cualquiera que sea el valor de x_1, tenemos que $S = 0$. Por tanto, para una muestra $n = 1$, no tenemos información sobre la variación. Podemos obtener información sobre la variación solamente cuando $n \geqq 2$, y n observaciones contienen información sobre $(n - 1)$ variables. Este número se llama el *grado de libertad.*

Representándolo por ϕ, en este caso tenemos

$$\phi = n - 1 \tag{9.11}$$

(4) Distribución de *u*

Este teorema se deriva fácilmente del teorema 9.2 y la estandarización de la variante normal.

> **Teorema 9.6**
>
> Sea $\bar{x}$ la media de una muestra $(x_1, \ldots, x_n)$ de $N(\mu, \sigma^2)$. Entonces la estadística
>
> $$u = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \tag{9.12}$$
>
> se distribuye como la distribución normal.

Designamos el punto del porcentaje α bilateral de la distribución normal por $u(\alpha)$. Es decir,

$$\Pr\{|u| \geqq u(\alpha)\} = \alpha, \tag{9.13}$$

donde u se distribuye como $N(0, 1^2)$.

La distribución u se usa para la prueba y estimación de la media de una población cuando se conoce α.



(5) Distribución de t

En las figuras 9.5 y 9.6 vemos que la distribución de la estadística t es similar a la distribución de u, pero tiene una variación ligeramente mayor.

Se ha establecido matemáticamente lo siguiente.

Teorema 9.7

Reemplazando σ por la desviación estándar de la muestra s en (9.12), tenemos

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} . \tag{9.14}$$

t se distribuye como la distribución t con un grado de libertad $\phi = n - 1$.

Debido a que se reemplaza σ por su estimador, es natural que la distribución t tenga una variación mayor que la distribución normal estándar. Cuando el grado de libertad $\phi = n - 1$ es pequeño, la distribución tiene extremos largos. Si n es muy grande, s se aproximará a σ. De manera que podemos creer que para n grande la distribución t diferirá poco de la distribución normal estándar. En realidad, una distribución t con un grado de libertad $\phi = \infty$ se conforma a la distribución normal.

La distribución t está claramente determinada por el grado de libertad ϕ, como se muestra en la figura 9.9. Designamos el punto del porcentaje bilateral α de la distribución t con grado de libertad ϕ por $t(\phi, \alpha)$, es decir,

$$\Pr\{\,|t| \geqq t(\phi, \alpha)\} = \alpha \tag{9.15}$$

donde t se distribuye como la distribución t con grado de liber-



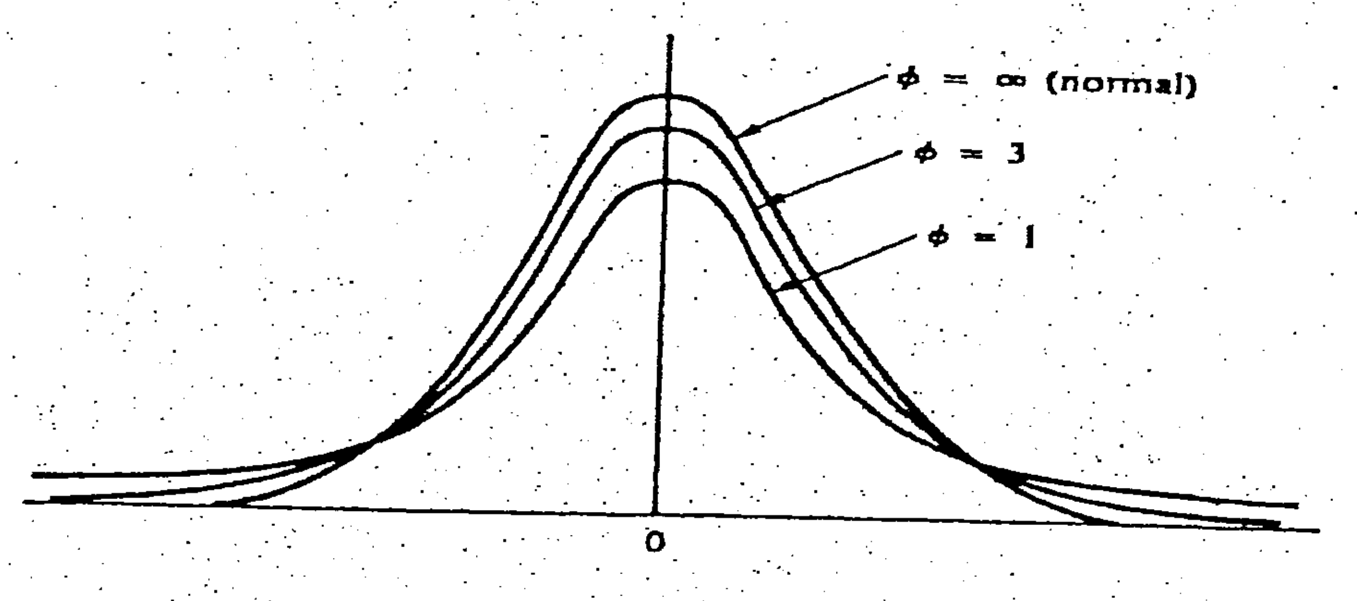


Figura 9.9 Distribución t

tad ϕ. La tabla A.3 del apéndice de este libro da los puntos del porcentaje.

La distribución t se usa para la prueba y estimación de la media de una población cuando se desconoce σ, o para someter a prueba y estimar la diferencia entre las medias de dos poblaciones.

9.3 PRUEBA DE LA HIPÓTESIS

Ejemplo 9.1

La resistencia a la tensión del acero inoxidable producido en cierta fábrica había sido estable, con una resistencia media de 72 kg/mm2 y una desviación estándar de 2.0 kg/mm^2. Recientemente, se ajustó una máquina. Para determinar el efecto del ajuste, se sometieron a prueba diez muestras, como se registra en la página 170.

Muestra No. i	Resistencia x_i (kg/mm²)
1	76.2
2	78.3
3	76.4
4	74.7
5	72.6
6	78.4
7	75.7
8	70.2
9	73.3
10	74.2

Suponga que la desviación estándar es la misma que se tenía antes del ajuste. ¿Podemos concluir que el ajuste ha cambiado la resistencia de tensión del acero?

Primero aclaremos el sentido de la pregunta:

1) Antes del ajuste, la resistencia de tensión del acero se distribuye como $N(\mu_0, \sigma_0^2)$. Aquí, μ_0 y σ_0 se conocen como $\mu_0 = 72.0$ y $\sigma_0 = 2.0$, respectivamente.
2) Después del ajuste, la resistencia se distribuye como $N(\mu, \sigma^2)$, donde se supone que σ es igual a $\sigma_0 = 2.0$, pero μ es desconocida.
3) La pregunta es si μ es igual a μ_0 o no. Es decir, éste es un problema de la comparación de las medias de dos poblaciones μ_0 y μ.
4) Diez muestras son muestras aleatorias de $N(\mu, \sigma^2)$. Ellas tienen alguna información sobre μ.

Tomando la media de $\bar{x}$ de las 10 muestras, tenemos

$$\bar{x} = \frac{76.2 + \cdots + 74.2}{10} = 75.0 \,. \qquad (9.16)$$

Este valor es diferente del de la media antes del ajuste, $\mu_0 = 72.0$ kg/mm². Sin embargo, no podemos concluir que el ajuste haya cambiado la resistencia del acero, porque la media de

la muestra $\bar{x}$ tiene variación y no es siempre igual a la media de la población.

Consideremos la hipótesis de que el ajuste no ha cambiado la resistencia del acero. Si esta hipótesis es verdadera, entonces $\bar{x}$ se distribuiría normalmente con una media $\mu_0 = 72.0$ y una desviación estándar

$$\sigma / \sqrt{n} = 2.0 / \sqrt{10} = 0.632\,. \tag{9.17}$$

Por tanto, la variable estandarizada μ_0 dada por

$$u_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 72.0}{0.632}$$

se distribuye como $N(0, 1^2)$. Puesto que tenemos $\bar{x} = 75.0$ de (9.16), el valor de μ_0 es

$$u_0 = \frac{75.0 - 72.0}{0.632} = 4.74\,. \tag{9.18}$$

La probabilidad de que $|\mu_0|$ alcance 4.74 es muy pequeña (= 0.000002). Esto quiere decir que ha sucedido algo muy improbable o que la hipótesis es incorrecta. Sospecharíamos por tanto que la hipótesis, $\mu = \mu_0$ (= 72.0), no es verdadera. Decidir a partir de una muestra de observaciones si una hipótesis sobre un parámetro de una población, tal como $\mu = \mu_0$ (= 72.0), es o no es correcta, se llama una *prueba de hipótesis*.

Cuando consideramos que la hipótesis no es correcta, decimos que la *rechazamos*. En el ejemplo, cuando la hipótesis H_0: $\mu = 72.0$ es verdadera, μ_0 tiende a ser cercana a cero. Por eso cuando $|\mu_0|$ es mayor que un cierto límite, rechazamos H_0. Por ejemplo, podemos seguir el procedimiento:

"Cuando $|\mu_0| \geqq 1.96$, rechace H_0, y cuando $|\mu_0| < 1.96$ acepte H_0".

Según esto, la probabilidad de rechazar H_0 cuando es verdadera es

$$\Pr(|u_0| \geqq 1.96) = 0.05. \qquad (9.19)$$

Esta probabilidad se llama un *nivel de significación* y por lo general se designa por α. Este tipo de error, en el cual una hipótesis se rechaza incorrectamente, se llama un *error de primer tipo.* El valor de α se selecciona generalmente en 0.05 (5%) o 0.01 (1%). La región de μ_0 en la cual la hipótesis H_0 se rechaza se llama la *región de rechazo,* y la región en la cual H_0 se acepta se llama la *región de aceptación.*

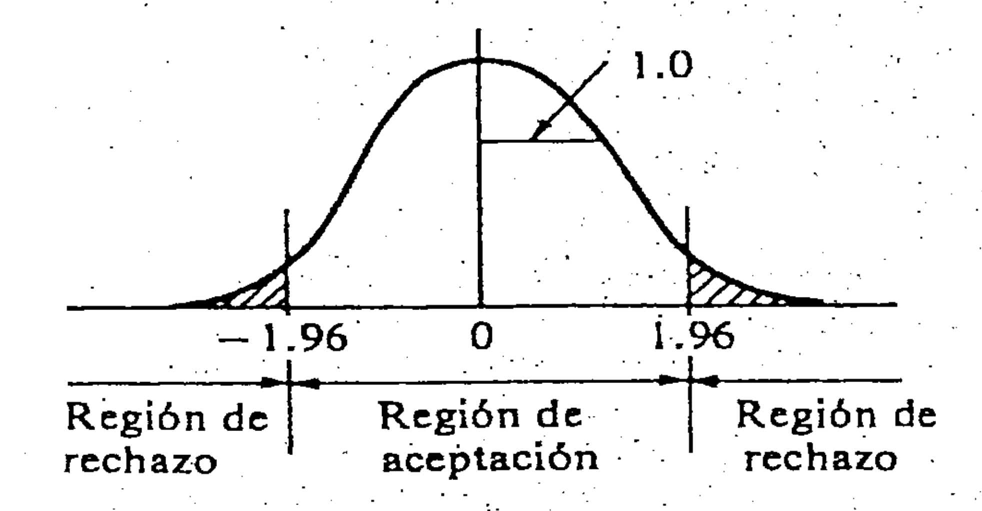


Figura 9.10 Regiones de rechazo y aceptación (prueba bilateral)

Cuando se rechaza una hipótesis, puede decirse con convicción que "el valor del parámetro especificado en la hipótesis no es correcto". Por otra parte, aun si aceptamos una hipótesis, no podemos concluir inequívocamente que la hipótesis sea correcta, porque podrían aceptarse muchas otras hipótesis respecto a una muestra de observaciones dada y, sin embargo, sólo una hipótesis puede ser verdadera. Cuando aceptamos una hipótesis que realmente no es verdadera, este error se llama *error de segundo tipo,* y su probabilidad se designa por β. Cuando $\mu \neq \mu_0$, escribiendo



$$u_1 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}, \tag{9.20}$$

La probabilidad de error de segundo tipo se ilustra en la figura 9.11. Cuando una hipótesis no es verdadera, muy probablemente la rechazaremos. La probabilidad de rechazar una hipótesis falsa se llama la *potencia de la prueba*, y se designa como P, donde $P = 1 - \beta$. El valor de P varía según la diferencia entre μ y μ_0, como se muestra en la figura 9.12.

El procedimiento para someter a prueba una hipótesis se muestra en la tabla 9.2.

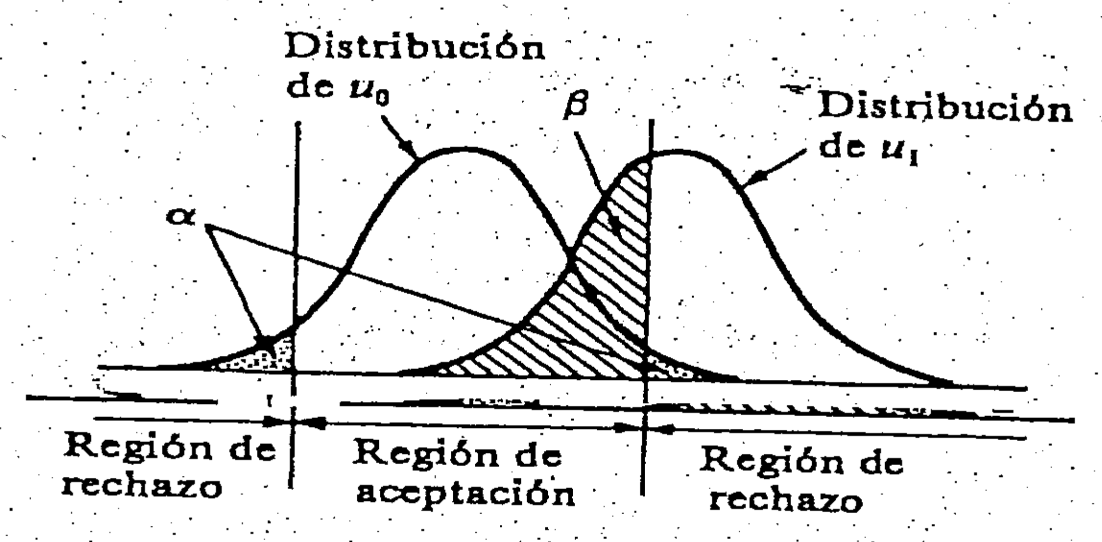


Figura 9.11 Error del segundo tipo, β

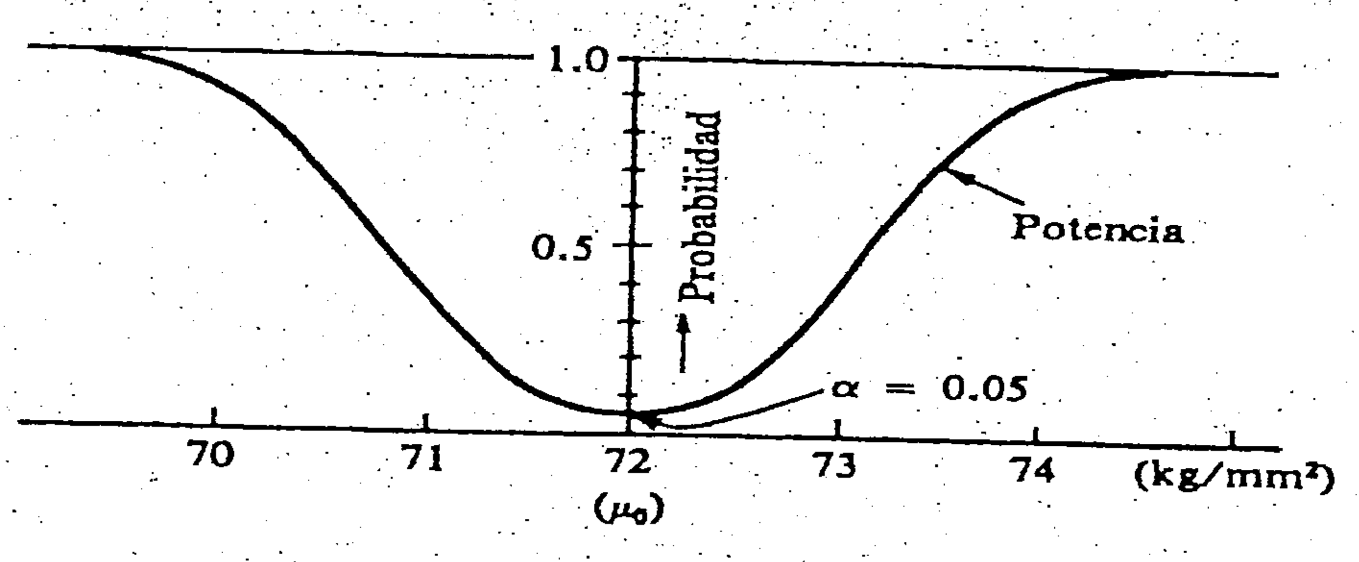


Figura 9.12 Curva de potencia para el ejemplo 9.1

Procedimiento	Ejemplo (*Ejemplo 9.1*)
1. Hipótesis, nivel de significación H_0: $\mu = \mu_0$ H_1 : $\mu \neq \mu_0$ (α = 0.05 o 0.01)	1. Hipótesis, nivel de significación H_0: $\mu = \mu_0$ (= 72.0) H_1 : $\mu \neq \mu_0$ (= 72.0)(α = 0.05)
2. Estadística $\bar{x} = \frac{1}{n}\sum x_i$ $u_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	2. Estadística $\bar{x} = \frac{1}{10}(76.2 + \cdots + 74.2) = 75.0$ $u_0 = \frac{75.0 - 72.0}{2/\sqrt{10}} = 4.74$
3. Prueba Obtenga u (α) de la tabla de la distribución normal. Si $\lvert u_0 \rvert \geqq u(\alpha) \rightarrow$ rechace H_0. Si $\lvert u_0 \rvert < u(\alpha) \rightarrow$ acepte H_0.	3. Prueba $u(0.05) = 1.96$ $\lvert u_0 \rvert = 4.74 > 1.96 = \mu(0.05)$ $\rightarrow$ rechace H_0.
4. Estimación $\hat{\mu} = \bar{x}$ $\mu = \bar{x} \pm u(\alpha)\frac{\sigma}{\sqrt{n}}$	4. Estimación $\hat{\mu} = \bar{x} = 75.0$ (kg/mm²) $\mu = 75.0 \pm 1.96 \times \frac{2.0}{\sqrt{10}}$ $= 75.0 \pm 1.24$ (kg/mm²)
5. Conclusión	5. Conclusión Debe concluirse, al nivel de significación del 5%, que el ajuste de la máquina ha cambiado la resistencia de tensión del acero. Los límites de confianza al 95% para la media son 75.0 ± 1.24 kg/mm².

Tabla 9.2 Prueba y estimación de la media de una población cuando σ es conocida (prueba de un solo lado)

Ejemplo 9.2

En el ejemplo 9.1, ¿podemos considerar que el ajuste de la máquina ha aumentado la resistencia del acero? Supongamos que las otras condiciones permanecen constantes.

En el ejemplo 9.1, nos interesaba saber si el ajuste de la máquina había cambiado la resistencia o no. Cuando rechazamos la hipótesis H_0: $\mu = 72.0$, consideramos que $\mu \neq 72.0$. En este ejemplo, nuestra pregunta es si el ajuste ha aumentado o no la resistencia. Si la hipótesis se rechazara, podríamos considerar que $\mu > 72.0$.

Cuando una hipótesis H_0 se rechaza, la hipótesis que aceptamos se llama ***hipótesis alternativa***, la cual se designa por H_1. En en el ejemplo 9.1, era

$$H_1: \mu \neq \mu_0 \quad (= 72.0) \tag{9.21}$$

mientras en el ejemplo 9.2, es

$$H_1: \mu > \mu_0 \quad (= 72.0). \tag{9.22}$$

La hipótesis H_0 se llama también ***hipótesis nula***, por oposición a la hipótesis alternativa.

Cuando se somete a prueba la hipótesis alternativa (9.21), la región de rechazo se encuentra en ambos lados de la distribución, como se muestra en la figura 9.10. Para (9.22), por otra parte, la región de rechazo se encuentra en el lado derecho, como se muestra en la figura 9.13. La primera de estas pruebas se llama ***prueba bilateral*** y la última de ellas se llama ***prueba unilateral***.

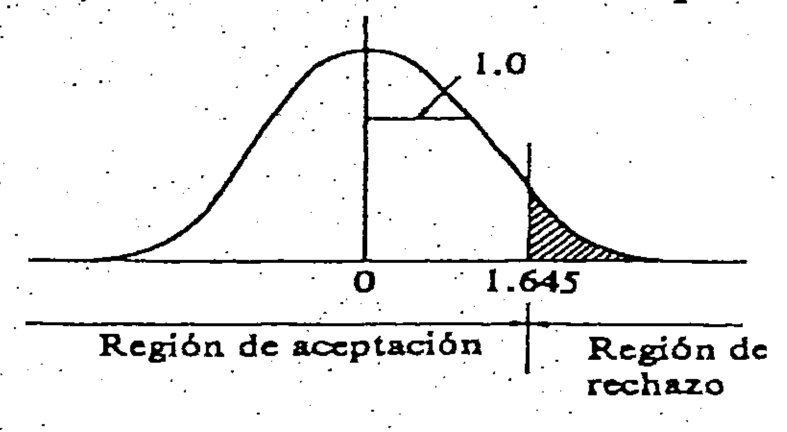


Figura 9.13 Regiones de aceptación y de rechazo (prueba unilateral)

El procedimiento para el ejemplo 9.2 se muestra en la tabla 9.3.

Procedimiento	Ejemplo (*Ejemplo 9.2*)
1. Hipótesis, nivel de significación H_0: $\mu = \mu_0$ $H_1: \mu > \mu_0 (\alpha = 0.05$ o $0.01)$ (o $H_1: \mu < \mu_0$)	1. Hipótesis, nivel de significación H_0: $\mu = \mu_0$ (= 72.0) $H_1: \mu > \mu_0$ (= 72.0)(α =0.05)
2. Estadística Como en el ejemplo 9.1	2. Estadística Como en el ejemplo 9.1
3. Prueba Obtenga u (2α) de la tabla de distribución normal. Cuando $H_1: \mu > \mu_0$ Si $\|u_0\| \geqq u(\alpha) \rightarrow$ rechace H_0. Si $\|u_0\| < u(\alpha) \rightarrow$ acepte H_0. Cuando $H_1: \mu < \mu_0$ Si $u_0 \leqq -u(2\alpha) \rightarrow$ rechace H_0. Si $u_0 > -u(2\alpha) \rightarrow$ acepte H_0.	3. Prueba $\mu(0.10) = 1.645$ $u_0 = 4.74 > 1.645 = \mu(0.10)$ $\rightarrow$ rechace H_0.
4. Estimación Como en el ejemplo 9.1	4. Estimación Como en el ejemplo 9.1
5. Conclusión	5. Conclusión Debe concluirse, a un nivel de significación del 5%, que el ajuste de la máquina ha aumentado la resistencia del acero. Los límites de confianza al 95% para la media son 75.0 ± 1.24 kg/mm².

Tabla 9.3 Prueba y estimación de la media de una población (prueba unilateral)



9.4 ESTIMACIÓN DE PARÁMETROS

Ejemplo 9.3

En el ejemplo 9.1, consideramos que la media de la resistencia había cambiado. Entonces, ¿cuál es el valor de la nueva media de la población?

La *estimación* es el proceso de análisis de una muestra con el fin de predecir el valor correspondiente del parámetro de la población. Un *estimador puntual* es un estimador de un parámetro de la población dado por un solo valor. El estimador puntual de un parámetro θ se designa por $\hat{\theta}$. Por ejemplo, la media de la población μ en el ejemplo 9.1 se estima por $\bar{x}$,

$$\hat{\mu} = \bar{x} = 75.0 \ (\text{kg/mm}^2) \tag{9.23}$$

Un *estimador de intervalo* es un estimador de un parámetro de una población dado por dos valores entre los cuales se considera que se encuentra el parámetro.

Suponiendo que $x_1, x_2, \ldots, x_n$, pertenecen a $N(\mu, \sigma^2)$, la media de la muestra $\bar{x}$ se distribuye como $N(\mu, \frac{\sigma^2}{n})$.

Escribiendo

$$u = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}, \tag{9.24}$$

la distribución de u es $N(0, 1^2)$. Por tanto, la probabilidad de que u se encuentre entre $\pm u(\alpha)$ es igual a $1 - \alpha$, donde $u(\alpha)$ designa el punto de porcentaje bilateral α de la distribución normal.

O,

$$\Pr\left\{-u(\alpha) < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < u(\alpha)\right\} = 1 - \alpha . \tag{9.25}$$

Reorganizando (9.25), tenemos

$$\Pr\left\{-u(\alpha) < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \mu(\alpha)\right\} = 1-\alpha. \tag{9.26}$$

Este intervalo se llama el *intervalo de confianza* 100 $(1-\alpha)$% para μ, y los límites superior e inferior del intervalo se llaman los *límites de confianza*. (9.26) significa que la probabilidad de que el intervalo contenga μ es $1-\alpha$. Esta probabilidad, es decir $(1-\alpha)$, se llama el *nivel de confianza*. Generalmente, el nivel de confianza se selecciona en 0.95 o en 0.99. Seleccionando 0.95, en (9.26), los límites de confianza del 95% son

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}. \tag{9.27}$$

La respuesta al ejemplo 9.3 se muestra en el paso 4 de la tabla 9.2.

9.5 PRUEBAS Y ESTIMACIONES DE LAS MEDIAS DE POBLACIONES CUANDO σ NO SE CONOCE

Según los problemas, pueden considerarse otras clases de pruebas y estimaciones. Los conceptos y procedimientos básicos son casi los mismos que se han descrito en las secciones previas, excepto por las estadísticas utilizadas. En las secciones siguientes se explicarán algunas pruebas y estimaciones comunes. En primer lugar, en esta sección discutiremos la prueba y estimación de la media de una población cuando σ es desconocida.

Supongamos que n muestras $x_1, x_2, \ldots, x_n$, pertenecen a una población con media μ y una desviación estándar σ. Entonces la distribución muestral de

$$u = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \tag{9.28}$$

puede considerarse como la distribución normal. Esto se aplica a la prueba y estimación de la media de una población cuando σ es conocida. Sin embargo, a veces ocurre que σ no se conoce. En este caso, parece natural reemplazar σ por s y usar

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \tag{9.29}$$

en lugar de u, donde s es la desviación estándar de la muestra,

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \tag{9.30}$$

Para usar esta estadística para la prueba y estimación, debemos conocer la distribución muestral de t. Esta distribución se conoce como la distribución t con $(n - 1)$ grados de libertad.

Para someter a prueba una hipótesis,

$$\begin{aligned} &H_0: \mu = \mu_0, \\ &H_1: \mu \neq \mu_0, \end{aligned} \tag{9.31}$$

calcule t_0 a partir de

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}, \tag{9.32}$$

y compárela con $t\,(n - 1, \alpha)$ donde $t\,(n - 1, \alpha)$ designa el punto de porcentaje de α bilateral de la distribución t con $(n - 1)$ grados de libertad. Cuando $|t_0| \geqq t\,(n - 1, \alpha)$, rechace H_0, y cuando



$|t_0| < t\,(n-1, \alpha)$, acepte H_0. En el caso de la prueba unilateral, es decir H_1: $\mu > \mu_0$ o H_1: $\mu < \mu_0$, t_0 se compara con t $(n-1, 2\alpha)$ o $-t$, $(n-1, 2\alpha)$, respectivamente.

Como el valor de t definido en (9.29) se distribuye como una distribución t con ϕ $n-1$, tenemos

$$\Pr\{-t\,(n-1, \alpha) < t < t\,(n-1, \alpha)\} = 1 - \alpha\,. \qquad (9.33)$$

Sustituyendo $\dfrac{\bar{x}-\mu}{s/\sqrt{n}}$ por t en (9.33) y reordenando, obtenemos

$$\Pr\left\{\bar{x} - t(n-1, \alpha)\frac{s}{\sqrt{n}} < u < \bar{x} + t(n-1, \alpha)\frac{s}{\sqrt{n}}\right\} = 1 - \alpha\,. \qquad (9.34)$$

Los límites de confianza de μ al 95% son

$$\bar{x} + t(n-1, 0.05)\frac{s}{\sqrt{n}}\,. \qquad (9.35)$$

Un ejemplo y el procedimiento para el análisis se muestran en la tabla 9.4.

Ejemplo 9.4

Suponga, en el ejemplo 9.1, que no puede asumirse que la desviación estándar de la resistencia después del ajuste sea la misma que era antes. ¿Podemos creer que el ajuste haya cambiado la resistencia del acero?

Procedimiento	Ejemplo
1. Hipótesis, nivel de significación H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$ ($\alpha = 0.05$ o 0.01)	1. Hipótesis, nivel de significación H_0: $\mu = \mu_0$ (= 72.0) H_1: $\mu \neq \mu_0$ ($\alpha = 0.05$)



2. Estadística

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$S = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$s = \sqrt{\frac{S}{n-1}}$$

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

3. Prueba
Obtenga el valor de t $(n-1, \alpha)$ de la tabla t.
Si $|t_0| \geqq t\ (n-1, \alpha) \rightarrow$ rechace H_0.
Si $|t_0| < t\ (n-1, \alpha) \rightarrow$ acepte H_0.

4. Estimación

$$\mu = 75.0 \pm 2.262 \times \frac{2.56}{\sqrt{10}}$$
$$= 75.0 \pm 1.83\ (\text{kg/mm}^2)$$

$$\hat{\mu} = \bar{x}$$

$$\mu = \bar{x} \pm t\,(n-1, \alpha)\frac{s}{\sqrt{n}}$$

5. Conclusión

2. Estadística

$$\sum x_i = 750.0$$
$$\sum x_i^2 = 56308.76$$
$$\bar{x} = 75.0$$

$$s = \sqrt{\frac{58.76}{9}} = \sqrt{6.53} = 2.56$$

$$S = 56308.76 - \frac{75.0^2}{10} = 58.76$$

$$t_0 = \frac{75.0 - 72.0}{2.56/\sqrt{10}} = 3.71$$

3. Prueba

$t(9, 0.05) = 2.262$
$|t_0| = 3.71 > 2.262$
$= t(9, 0.05) \rightarrow$ rechace H_0.

4. Estimación

$$\hat{\mu} = 75.0$$

5. Conclusión
Debe concluirse, al nivel de significación del 5%, que el ajuste de la máquina ha cambiado la resistencia del acero. Los límites de confianza para la media al nivel del 95% son 75.0 ± 1.83 kg/mm².

Tabla 9.4 Prueba y estimación de la media de una población cuando σ es desconocida

9.6 PRUEBAS Y ESTIMACIONES DE LAS DIFERENCIAS ENTRE LAS MEDIAS DE DOS POBLACIONES

Supongamos que se tienen dos conjuntos de muestras y queremos saber si las medias de dos poblaciones son iguales o no.

Sean $x_{11}, x_{12}, \ldots, x_{1n1}$, para n_1 muestras de la primera población con una media μ_1 y desviación estándar σ, y $x_{21}, x_{22}, \ldots, x_{2n2}$, para n_2 muestras de la segunda población con una media μ_2 y desviación estándar σ. Para someter a prueba la diferencia entre las medias de dos poblaciones, examinamos la diferencia $\bar{x}_1 - \bar{x}_2$. Debido a que las distribuciones de las medias de las muestras $\bar{x}_1$ y $\bar{x}_2$ pueden considerarse como $N(\mu_1, \frac{\sigma^2}{n_1})$ y $N(\mu_2, \frac{\sigma^2}{n_2})$, respectivamente, $\bar{x}_1 - \bar{x}_2$ tiene una distribución normal,

$$N(\mu_1 - \mu_2, (\frac{1}{n_1} + \frac{1}{n_2})\sigma^2) .$$

Por tanto, estandarizando esto,

$$u = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{9.36}$$

se distribuye como $N(0, 1^2)$. Supongamos ahora que σ es desconocida. Reemplazando σ por s en (9.36), tenemos

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{9.37}$$

donde

$$s = \sqrt{\frac{S_1 + S_2}{n_1 + n_2 - 2}} \tag{9.38}$$

y S_1 o S_2 es la suma de los cuadrados en cada muestra. En (9.37), t tiene una distribución t con $\phi = n_1 + n_2 - 2$. Si no hay diferencia entre las dos medias, colocando $\mu_1 = \mu_2$ en (9.37), tenemos

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{9.39}$$

Un ejemplo del procedimiento para el análisis se muestra en la tabla 9.5.

Ejemplo 9.5.

Se estudió el efecto de la corriente de soldadura sobre la fuerza de ruptura.

x_1 (600A)	37	29	35	28	24	36	40	37	33	28	29		
x_2 (800A)	22	32	27	30	24	34	32	20	24	25	28	26	26

Tabla 9.5 Prueba y estimación de la diferencia entre las medias de dos poblaciones

¿Hay alguna diferencia en la fuerza de ruptura de las soldaduras usando estas dos corrientes?



Procedimiento	Ejemplo
1. Hipótesis, nivel de significación H_0: $\mu_1 = \mu_2$ H_1 : $\mu_1 \neq \mu_2$ ($\alpha = 0.05$ o 0.01)	1. Hipótesis, nivel de significación H_0: $\mu_1 = \mu_2$ H_1 : $\mu_1 \neq \mu_2$ ($\alpha = 0.05$)
2. Estadística $\bar{x}_1 = \frac{1}{n_1} \sum x_{1i}$ $\bar{x}_2 = \frac{1}{n_2} \sum x_{2i}$ $S_1 = \sum x_{1i}^2 - \frac{(\sum x_{1i})^2}{n_1}$ $S_2 = \sum x_{2i}^2 - \frac{(\sum x_{2i})^2}{n2}$ $s = \sqrt{\frac{S_1 + S_2}{n_1 + n_2 - 2}}$ $t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	2. Estadística $\sum x_{1i} = 366$ $\sum x_{1i}^2 = 12454$ $\sum x_{2i} = 350$ $\sum x_{2i}^2 = 9630$ $\bar{x}_1 = \frac{366}{11} = 33.27$ $\bar{x}_2 = \frac{350}{13} = 26.92$ $S_1 = 12454 - \frac{366^2}{11} = 276.18$ $S_2 = 9630 - \frac{350^2}{13} = 206.92$ $s = \sqrt{\frac{276.18 + 206.92}{11 + 13 - 2}} = \sqrt{21.96} = 4.69$ $t_0 = \frac{33.27 - 26.92}{4.69\sqrt{\frac{1}{11} + \frac{1}{13}}} = 3.30$
3. Prueba Obtenga el valor de t $(n_1 + n_2 - 2, \alpha)$ de la tabla t. Si $\|t_0\| \geqq t\,(n_1 + n_2 - 2, \alpha) \rightarrow$ rechace H_0.	3. Prueba $t\,(22, 0.05) = 2.074$ $\|t_0\| = 3.30 > 2.074$ $= t\,(22, 0.05) \rightarrow$ rechace H_0.

Si $|t_0| < t(n_1 + n_2 - 2, \alpha) \rightarrow$
acepte H_0.

4. Estimación

$$\widehat{\mu_1 - \mu_2} = \bar{x}_1 - \bar{x}_2$$

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2$$

$$\pm t(n_1 + n_2 - 2) \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

5. Conclusión

4. Estimación

$$\widehat{\mu_1 - \mu_2} = 33.27 - 26.92 = 6.35$$

$$\mu_1 - \mu_2 = 6.35 \pm 2.074 \times 4.69 \times \frac{1}{11} + \frac{1}{13}$$

$$= 6.35 \pm 3.98 \text{ (kg/cm)}$$

5. Conclusión

Debe concluirse que, al nivel de significación del 5%, hay una diferencia entre las fuerzas de ruptura producidas por las dos corrientes de soldadura. Los límites de confianza del 95% para la diferencia entre las dos medias son 6.35 ± 3.98 kg/cm.

9.7 PRUEBAS Y ESTIMACIONES EN OBSERVACIONES PAREADAS

Cuando se toman muestras de dos poblaciones, a veces resultan diferencias significativas causadas por factores extraños, sin que exista diferencia en el efecto que estamos tratando de medir. Por ejemplo, en un experimento para ver cuál de dos abonos (A o B) es mejor, se sembraron dos parcelas de trigo en cada una de diez estaciones experimentales. En cada estación, una de las parcelas se abonó con el abono A y la otra con el abono B. Si el promedio de las 10 observaciones relativas al abono A se compara con el promedio de las 10 observaciones relativas al abono B, la dife-

rencia observada (si la hay) puede deberse a las diferencias en el tipo de suelo o a diversas condiciones climáticas, en lugar de ser causada por diferencias entre los abonos. También existe la posibilidad de que éstos causen una diferencia, pero que ella se vea oscurecida por otros factores. Un tipo de experimento que puede superar estas dificultades consiste en que las observaciones se hagan por pares. Tratamos de asegurarnos de que los dos miembros de cada uno de los pares sean semejantes en todo, excepto en aquello que deseamos medir. Así, cada par de parcelas tendría aproximadamente el mismo tipo de suelo, condiciones climáticas, etc.

Sea x_{1i} el primer miembro del i-ésimo par y x_{2i} el segundo. Supongamos que tenemos n pares de observaciones,

$$(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1i}, x_{2i}), \ldots, (x_{1n}, x_{2n}) \qquad (9.40)$$

Si tomamos las diferencias $d_i = x_{1i} - x_{2i}$, obtenemos una serie de n observaciones d_i. Deseamos someter a prueba la hipótesis $\mu_1 = \mu_2$. Esta hipótesis afirma que no hay diferencia entre los tratamientos, es decir, que no hay diferencia dentro de los pares. Puede haber una diferencia entre los pares, pero ésta puede eliminarse tomando la diferencia d_i. Si la hipótesis es verdadera, los valores de d_i provendrían de una población con media cero. La prueba se realiza exactamente de la misma manera indicada en la sección 9.5 (prueba t con $\phi = n - 1$).

Un ejemplo y el procedimiento para el análisis se muestran en la tabla 9.6.

Ejemplo 9.6

Se compararon las longitudes de 10 muestras antes y después de haber sido sometidas a horneado ultrasónico. ¿El tratamiento por calor causa un cambio en la longitud?



Muestra	Antes	Después (mm)
1	11.94	12.00
2	11.99	11.99
3	11.98	11.95
4	12.03	12.07
5	12.03	12.03
6	11.96	11.98
7	11.95	12.03
8	11.96	12.02
9	11.92	12.01
10	12.00	11.99

Procedimiento

1. Hipótesis, nivel de significación

H_0: $\mu_1 - \mu_2 = 0$

H_1 : $\mu_1 - \mu_2 \neq 0$

($\alpha = 0.05$ o 0.01)

2. Estadística

$$d_i = x_{1i} - x_{2i}$$

$$\bar{d} = \frac{1}{n}\sum d_i$$

$$s = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}}$$

$$t_0 = \frac{\bar{d}}{s/\sqrt{n}}$$

Ejemplo

1. Hipótesis, nivel de significación

H_0: $\mu_1 - \mu_2 = 0$

H_1 : $\mu_1 - \mu_2 \neq 0$ ($\alpha = 0.05$)

2. Estadística

x_{1i}	x_{2i}	d_i	d_i^2
11.94	12.00	−0.06	0.0036
11.99	11.99	0.00	0.0000
11.98	11.95	0.03	0.0009
12.03	12.07	−0.04	0.0016
12.03	12.03	0.00	0.0000
11.96	11.98	−0.02	0.0004
11.95	12.03	−0.08	0.0064
11.96	12.02	−0.06	0.0036
11.92	12.01	−0.09	0.0081
12.00	11.99	0.01	0.0001
Total		−0.31	0.0247



$$\overline{d} = \frac{-0.31}{10} = -0.031$$

$$s = \sqrt{\frac{0.0247 - \frac{(-0.31)^2}{10}}{9}}$$

$$= \sqrt{0.0016766} = 0.0409$$

$$t_0 = \frac{-0.031}{0.0409/\sqrt{10}} = -2.40$$

3. Prueba
Obtenga $t\ (n-1, \alpha)$ en la tabla t.
Si $|t_0| \geqq t\ (n-1, \alpha) \rightarrow$ rechace H_0.
Si $|t_0| < t\ (n-1, \alpha) \rightarrow$ acepte H_0.

3. Prueba
$t(9, 0.05) = 2.262$
$|t_0| = 2.40 > 2.262$
$= t(9, 0.05) \rightarrow$ rechace H_0.

4. Estimación

$$\widehat{\mu_1 - \mu_2} = \overline{d}$$

$$\mu_1 - \mu_2 = \overline{d} \pm t(n-1, \alpha)\frac{s}{\sqrt{n}}$$

4. Estimación

$$\widehat{\mu_1 - \mu_2} = -0.031$$

$$\mu_1 - \mu_2 = -0.031 \pm 2.262 \times \frac{0.0409}{\sqrt{10}}$$

$$= -0.031 \pm 0.029 (\text{mm})$$

5. Conclusión

5. Conclusión
Debe concluirse, al nivel de significación del 5%, que el horneado ultrasónico causa un cambio en la longitud. Los límites de confianza para la diferencia en las medias son -0.031 ± 0.029 mm.

Tabla 9.6 Prueba y estimación en observaciones pareadas

9.8 PRUEBAS DE SIGNIFICACIÓN DE LOS COEFICIENTES DE CORRELACIÓN

Como ya se dijo en el capítulo 6, se calcula el coeficiente de correlación r cuando deseamos conocer la fuerza de la relación entre dos variables x y y. Aquí supongamos que x y y siguen una distribución normal bivariada con un coeficiente de correlación de la población igual a ρ. Aun cuando $\rho = 0$, el coeficiente de correlación de la muestra r no siempre es igual a cero. Se puede realizar la prueba para comprobar si el coeficiente ρ de la población es igual a cero o no, aplicando el siguiente teorema.

Teorema 9.8

Sea (x_i, y_i), $i = 1, 2, \ldots, n$ sea n muestras de una distribución normal bivariada con un coeficiente de correlación cero. Designando el coeficiente de correlación de la muestra con r,

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (9.41)$$

se distribuye como la distribución t con $(n - 2)$ grados de libertad.

Un ejemplo y el procedimiento para someter a prueba la hipótesis de que el coeficiente de correlación de una población es igual a cero se muestran en la tabla 9.7.

Ejemplo 9.7

En la tabla 6.1 (capítulo VI, p. 71) hay 30 pares de datos sobre la presión del aire de soplado y el porcentaje de tanques plásticos defectuosos. ¿Podemos concluir que existe una relación entre estas dos características?

Ejercicio 9.2

Llene los espacios en blanco en la tabla 9.1.

Ejercicio 9.3

Se requería un experimento para determinar si un cierto tratamiento de superficie aumentaba la resistencia a la abrasión de un tipo particular de material. Se tomaron diez piezas de prueba de material; se trataron cinco de ellas, y las otras cinco se dejaron sin tratamiento. Luego se midió la resistencia de abrasión de las diez piezas. En la tabla a continuación se muestran los resultados:

Tratado	Sin tratamiento
18.2	12.9
16.0	11.3
12.2	13.2
16.7	16.5
14.4	14.2

Responda las siguientes preguntas:

1) ¿Podemos concluir que el tratamiento aumentó la resistencia de abrasión del material? Realice una prueba de hipótesis.
2) Encuentre los límites de confianza al 95% para la diferencia en la media de la resistencia de abrasión entre los materiales tratados y los no tratados.
3) Si es posible, mejore el diseño del experimento anterior.



X

La ruta de la calidad

La ruta de la calidad es un procedimiento para solucionar problemas. En los términos usados en ella, un problema se define de la siguiente manera:

"Un problema es el resultado no deseado de una tarea".

La solución para un problema es mejorar el resultado deficiente hasta lograr un nivel razonable. Las causas de los problemas se investigan desde el punto de vista de los hechos, y se analiza con precisión la relación entre la causa y el efecto. Se evitan estrictamente las decisiones sin fundamento basadas en la imaginación o en la teoría desde un escritorio, debido a que los intentos de solucionar los problemas con base en esas decisiones orientan en direcciones equivocadas, lo cual lleva al fracaso o a demorar la mejoría. Se diseñan y se implementan medidas que contrarresten el problema para evitar que los factores causales vuelvan a presentarse. Este procedimiento es una especie de recuento o representación de las actividades de control de calidad, y por eso la gente lo llama la ruta de la calidad.

Un problema se soluciona de acuerdo con los siguientes siete pasos:



1) *Problema:*
Identificación del problema.

2) *Observación:*
Reconocimiento de las características del problema.

3) *Análisis:*
Búsqueda de las principales causas.

4) *Acción:*
Acción para eliminar las causas.

5) *Verificación:*
Confirmación de la efectividad de la acción.

6) *Estandarización:*
Eliminación permanente de las causas.

7) *Conclusión:*
Revisión de las actividades y planeación del trabajo futuro.

Si estos siete pasos se clarifican y se implementan en el mismo orden, las actividades de mejoría serán lógicamente consistentes y se acumularán establemente. A veces este procedimiento parece ser una forma larga de solucionar un problema, pero a la larga es el camino más corto y el más seguro.

Se describirá en detalle cada paso de la ruta de la calidad. Cada paso contiene varias "actividades" y el contenido de éstas se explicará en "notas".

10.1 EL PROBLEMA

Defina el problema con claridad

Actividades

1) Muestre que el problema que se está tratando es mucho más importante que cualquier otro.



2) Muestre cuál es el contexto del problema y qué curso ha seguido hasta el momento.
3) Exprese en términos concretos solamente los resultados no deseados del desempeño deficiente. Demuestre cuál es la pérdida en el actual desempeño y cuánto necesita mejorarse.
4) Plantee un tema y una meta y, si es necesario, subtemas.
5) Proponga a una persona para que se haga cargo de la tarea oficialmente. Cuando la tarea va a ser realizada por un equipo, nombre a sus miembros y al líder.
6) Presente un presupuesto estimado para la mejoría.
7) Haga un cronograma para la mejoría.

Notas

1) Nos rodean innumerables problemas, grandes y pequeños. Con personal, tiempo y dinero limitados, tenemos que asignar prioridades para seleccionar problemas. Utilice toda la información posible para identificar el problema más importante. Cuando usted selecciona un problema entre varios otros, debe estar seguro de las razones para esa selección.
2) Algunos problemas se seleccionan por su contexto o por el curso que han seguido hasta el momento. En estos casos, deben identificarse claramente las circunstancias. También aquí, se debe utilizar tanta información como sea posible. Dar las razones para solucionar un problema en particular no tiene influencia directa sobre la solución del mismo, pero es importante en un sentido indirecto. El paso que aclara el grado de importancia es necesario. Si el nivel de importancia es muy alto y muchas personas lo comprenden, el problema será tratado con seriedad. Esto hace que el problema tenga una alta posibilidad de ser solucionado. Por otra parte, cuando la gente no comprende completamente la importancia de un problema, aun de uno de fácil solución, sus esfuerzos serán apenas medianos cuando se les pide que lo solucionen, o abandonarán la tarea sin concluirla. Nada mejorará. Con el



fin de evitar esto, utilice diagramas y fotografías para explicar el desempeño inadecuado.

3) Sería un salto infundado tratar de describir factores causales y realizar acciones correctivas para esos factores mientas se esté todavía en la etapa de determinar cuál es el problema (tema). En esta etapa no se determinan factores causales, sino más tarde en la etapa de *análisis*. Se expresan solamente los resultados del problema, y se deben expresar correctamente con el fin de aclarar el tema. Después, se describen la pérdida de desempeño en la actual situación y las ventajas de lograr mejorías. Estos pasos se deben realizar en orden con el fin de ganar reconocimiento del problema de tanta gente como sea posible.

4) Es importante indicar las bases sobre las cuales se fijan los valores deseados. No se propondrán metas absurdas. El valor deseado podría ser una fracción de 0% de ítems defectuosos, pero, en la mayoría de los casos, este tipo de valores son objetivos ideales. Es muy difícil lograr este tipo de objetivos y aun si esos objetivos se pudieran lograr, al hacerlo se presentarían otros problemas. Debe fijarse como meta un valor razonable teniendo en cuenta la eficiencia y las posibilidades técnicas. Cuando el tema incluye muchas clases de problemas, divídalo en subtemas para poder enfrentar el problema adecuadamente. En algunos casos, cuando la totalidad consiste en muchas partes similares, sería mejor seleccionar una parte típica del todo para analizarla y usarla como base para ampliar hacia afuera hasta abarcar el todo. Una parte se selecciona y se usa como tema principal y la otra se usa como subtema.

5) Fije una fecha límite para lograr una solución del problema. Por lo general, si se comprende bien la necesidad, la cuestión de si se debe solucionar el problema también será clara. Independientemente de la magnitud del valor estimado del efecto, un problema que carece de un cronograma claramente definido será un problema con una baja prioridad.



10.2 OBSERVACIÓN

Investigue las características específicas del problema desde una amplia gama de puntos de vista

Actividades

1) Investigue cuatro puntos (tiempo, lugar, tipo y síntoma) para descubrir las características del problema.
2) Después investigue desde muchos puntos de vista para descubrir la variación en el resultado.
3) Vaya al lugar y recoja la información necesaria que no puede ponerse en forma de datos.

Notas

Investigue el problema desde puntos de vista diferentes y compréndalo plenamente en todos sus aspectos. En este paso, no se refiera a las causas del problema, sino limítese a mirarlo como es. A primera vista, esto es similar al paso anterior, *problema*. Estos dos pasos con frecuencia tienden a confundirse, pero su propósito es totalmente diferente. El objetivo del paso *problema* es reconocer su importancia; el objetivo del paso *observación* es descubrir los factores que son causas del problema. A veces se usa la misma información en los dos pasos, pero se usa con objetivos diferentes. Los investigadores criminales diestros y los detectives privados siempre usan una técnica común: ellos investigan minuciosamente el lugar del crimen antes de hacer otra cosa. Del lugar obtienen pistas sobre las cuales basar la búsqueda del autor, y gradualmente cierran el cerco alrededor del sospechoso. Si el investigador no estudia plenamente la situación en la cual se cometió el crimen antes de iniciar la búsqueda, no solamente no encontrará a la persona correcta, sino que puede terminar arres-

tando a otra completamente inocente. Lo mismo puede decirse de la solución de problemas.

1) Las pistas para la solución de un problema se encuentran en el problema mismo. Pueden descubrirse varios fenómenos en los resultados cuando un problema se observa desde diversos puntos de vista. Éstas son las características especiales del problema y son las pistas para su solución. La razón es que si hay una variación en los resultados también debe haberla en los factores causales, y es por tanto posible correlacionar los dos tipos de variación. La utilización de la variación en los resultados para descubrir la variación en factores causales es por tanto una manera efectiva de identificar los factores principales.
El mejor ángulo para observar un problema variará en cada caso. Pero, independientemente de cuál es el problema, hay por lo menos cuatro puntos de vista importantes desde los cuales se debe mirar. Éstos son tiempo, lugar, tipo y síntoma. A continuación vemos un ejemplo usado para mejorar la fracción de ítems defectuosos de cierto producto.

a) Investigue de la siguiente manera:
— ¿Hay diferencia en la fracción de defectuosos entre la mañana, el mediodía y la tarde?
— ¿Hay alguna diferencia en la fracción de defectuosos entre el lunes y el sábado?
También podríamos usar diversas escalas de tiempo, tales como las diferencias entre una y otra semana, ente un mes y otro, o durante diferentes estaciones, períodos, años.
b) En seguida, investigue desde el punto de vista de la localización del defecto en el producto.
— ¿Hay alguna diferencia en la fracción de defectuosos entre los paneles superiores, laterales o inferiores?
— Hay alguna diferencia en la fracción de defectuosos entre la localización de los productos en el horno (cer-

ca de la puerta, cerca de las ventanas, cerca de las paredes, en el centro del horno)?

También podríamos preguntar desde el punto de vista de la orientación (este, oeste, norte, sur) o de la altura (arriba y abajo). Cuando los productos son muy largos, ¿el defecto se presenta en la parte delantera, en la mitad o en la parte de atrás? ¿El defecto se presenta en las secciones rectas o curvas de un producto que tiene una forma complicada, o depende del área, es decir, área A, área B, etc.?

c) Haga una verificación con respecto al tipo.
— ¿Hay diferencias en la fracción de productos defectuosos de diferentes tipos hechos por la misma empresa?
— ¿Hay alguna diferencia en la fracción de productos defectuosos de un tipo similar encontrados en el pasado?

Podemos pensar en otros ángulos respecto al tipo, tales como especificaciones, clase, si lo usan niños o adultos, hombres o mujeres, o si el producto es para el mercado interno o para la exportación.

d) Finalmente, investigue desde el punto de vista del síntoma. Como ejemplo de síntomas en un ítem defectuoso, usemos la porosidad.
— ¿Hay diferencias en la forma de las porosidades observadas (son circulares, elípticas, angulares o de alguna otra forma)?
— ¿Hay alguna diferencia en la manera en la cual aparecen porosidades múltiples (rectas, curvas, continuas, discontinuas, etc.)?

Además, ¿de qué tamaño son las porosidades, en qué condiciones? ¿Aparecen en ciertas superficies (están en todas partes, o solamente en algunas áreas)? ¿Cuáles son las características del área circundante (cambios en color o en calidad, o aparición de objetos extraños)?

2) Cualquiera que sea el problema, la investigación debe hacerse, por lo menos, desde estas cuatro perspectivas. Sin embargo, por sí solas, ellas no son suficientes. El problema debe investigarse desde varios puntos de vista basados en las características del problema. Mientras más amplia sea la variación de los resultados descubiertos, mejor.
3) Generalmente, la solución de problemas debe basarse en datos. La información que no se basa en datos, es decir, la que se basa en la memoria o en la imaginación, puede usarse solamente como punto de referencia. Sin embargo, la información que no se puede obtener de datos a veces tiene un papel importante en la solución de los problemas. Dentro de lo posible, quienes participan en la investigación deben estar presentes en el lugar de los hechos, no en una oficina, sino en el lugar. Aquí ellos pueden observar y obtener información que no se puede poner en forma de datos. Este tipo de información, que funciona como un catalizador en una reacción química, da nuevas pistas durante el proceso de pensamiento para solucionar el problema.

10.3 ANÁLISIS

Descubra cuáles son las principales causas

Actividades

1) Plantee hipótesis de causas (seleccionando los candidatos más importantes como causas).

   a) Haga un diagrama de causa-efecto (que contiene todos los elementos que parecen estar relacionados con el problema) con el fin de recoger todo el conocimiento posible sobre las posibles causas.
   b) Utilice la información obtenida en el paso de *observación* y excluya todos los elementos que no son claramente re-



levantes. Revise el diagrama de causa-efecto usando los elementos restantes.

c) En el último diagrama, marque aquellos elementos que parecen tener una alta posibilidad de ser las principales causas.

2) Someta a prueba las hipótesis (deduzca las principales causas entre las señaladas).

a) Empezando por los elementos que tienen una alta posibilidad de ser causas, diseñe nuevos planes para verificar el efecto de esos elementos sobre el problema, bien sea obteniendo nueva información o realizando experimentos.

b) Integre toda la información investigada y concluya cuáles son las principales causas posibles.

c) Si es posible, reproduzca el problema intencionalmente.

Notas

Este paso se divide en dos partes; el primero es el planteamiento de las hipótesis, y el segundo, el proceso de someter a prueba esas hipótesis. La razón para esos pasos es que, en la ruta de la calidad, las causas deben determinarse científicamente. En muchos casos, la causa del problema se determina por medio de la discusión entre los interesados en solucionar el problema, o por medio de la decisión arbitraria de una persona. Las decisiones de este tipo son con frecuencia equivocadas, y la mayoría de las equivocaciones ocurren cuando se omite el paso de someter a prueba las hipótesis. Cuando pensamos en las causas (hipótesis), se discuten las razones y se analizan los datos. Es fácil incurrir en la ilusión de pensar que durante las discusiones y el análisis de la información la discusión y los datos verifican la corrección de una hipótesis. Pero el planteamiento de una hipótesis y el proceso de someterla a prueba son cosas diferentes, y no se pueden usar los mismos datos para ambos. Someter a prueba las hipótesis requiere nuevos datos que no se usan para plantearlas. La re-



colección de datos para someter a prueba las hipótesis debe planearse en forma lógica y las hipótesis deben someterse a prueba por medio de procedimientos estadísticos.

1) Para plantear las hipótesis, un diagrama de causa-efecto es una herramienta útil. Todos los elementos del diagrama son causas hipotéticas del problema. El diagrama debe contener los elementos que finalmente se identificarán como las principales causas.

   a) La expresión del efecto en el diagrama debe ser tan concreta como sea posible, pues si se expresa en términos abstractos, el número de elementos en el diagrama será enorme. Sin embargo, como una definición abstracta es una integración de varios casos individuales, cada uno de esos casos tiene partes que son innecesarias. Por ejemplo, si expresamos el efecto como un tipo de defecto, las causas en el diagrama serán una colección de factores que causan el defecto. Sin embargo, si los defectuosos son el efecto, incluyendo muchos tipos de defectos, es necesario recoger muchos tipos de defectos y el contenido del diagrama se diversificará. Así, mientras más concreta sea la expresión de las características, más efectivo será el diagrama.
   Primero, haga un diagrama de causa-efecto que tenga suficientes elementos para incluir todas las opiniones de quienes están involucrados en la solución del problema.
   b) No sería efectivo investigar todas las causas posibles, de manera que en este punto tenemos que reducir el número basándonos en la información. La información examinada en el paso de *observación* será muy útil aquí. Los elementos que no corresponden a la variación de los resultados se excluyen del diagrama.
   Digamos, por ejemplo, que la fracción de defectuosos es alta por la mañana y baja por la tarde. Si los trabajadores son los mismos en la mañana y en la tarde, eliminamos

los trabajadores del diagrama, puesto que no encajan con los resultados. Pero si en la mañana y en la tarde se usa diferente maquinaria, dejamos la maquinaria en el diagrama puesto que ella encaja con el resultado.

Si se han examinado varios resultados dispersos en el paso de *observación*, podemos eliminar del diagrama muchas de las posibles causas. Luego que se han eliminado en esta forma los elementos que no pueden ser causa, hacemos otro diagrama de causa-efecto utilizando los restantes elementos. El diagrama mejorará en la medida en que sea más pequeño (es decir, cuando sea menor el número de elementos).

c) No todos los elementos en el diagrama revisado tienen la misma posibilidad de causar el problema. Los elementos deben ordenarse por rango según sus probabilidades con base en la información obtenida en el paso de la *observación* y examinados en ese orden.

De esta manera hemos limitado los principales candidatos a causas, pero debemos tener presente que los candidatos todavía son solamente candidatos. En este momento, todavía no podemos determinar que los elementos evaluados como altamente posibles sean las principales causas del problema, debido a que hasta el momento se han usado los datos obtenidos para plantear las hipótesis, y es necesario usar datos obtenidos de acuerdo con un nuevo plan para decidir si las hipótesis son correctas o no.

2) El proceso de someter a prueba las hipótesis también debe basarse en datos obtenidos de experimentos y encuestas. La información se debe recoger de acuerdo con un plan cuidadosamente diseñado.

a) Someter a prueba la hipótesis es investigar si realmente existe una relación entre las posibles causas y los resultados, y, en caso de que exista, qué tan fuerte es la relación,

es decir, qué efecto tiene la posible causa. Hay varios métodos para expresar la fuerza de esa relación, como por ejemplo los coeficientes de correlación, el análisis de varianza y el diagrama de Pareto de causas. O, sencillamente, pueden marcarse en el diagrama de causa-efecto. Debemos evitar tomar decisiones sobre las principales causas por medio de "votos". La determinación de la principal causa por medio del voto es un método democrático, pero no hay ninguna garantía de que sea científicamente correcto. Hay muchos ejemplos de ocasiones en las cuales se ha seleccionado como causa un elemento por acuerdo unánime, y después de la investigación se encontró que el elemento no era una causa.

A veces, se implementan acciones remediales sin ningún análisis de datos. Se ponen en práctica todas las acciones que parecen prometer efectividad. Si se obtienen buenos resultados, se considera terminado el proceso de solución del problema. El orden seguido es exactamente el contrario del que aquí se ha defendido, porque lo que se está haciendo es investigar la causa por la acción. Para la solución de un problema, este procedimiento requiere de una gran cantidad de ensayos y errores. Si el problema se soluciona y aun si vemos que las acciones remediales son efectivas en la solución del problema, en la mayoría de los casos no podremos averiguar las verdaderas causas principales, porque la relación entre causas y acciones remediales no corresponde uno a uno.

b) La causa principal es uno o varios elementos que tienen la mayor influencia en los resultados. Muchos elementos pueden influir sobre el resultado en una u otra forma, de manera notable o no, pero sería inefectivo adoptar medidas para todos los elementos. Las acciones correctivas deben tomarse respecto a factores de gran causalidad y no respecto a aquéllos que tienen poco efecto. Ésta es la razón por la cual debemos investigar e integrar toda clase de información y decidir cuáles son las principales causas.



c) Puede encontrarse evidencia para la causa por medio de la reproducción intencional del defecto. Sin embargo, esta reproducción debe emprenderse con cuidado. Si usamos una unidad no estándar de algún producto, puede producirse un producto defectuoso, pero eso no quiere decir necesariamente que la unidad no estándar sea la causa del defecto. Otros factores pueden ser la causa del defecto. Un defecto producido intencionalmente debe tener las mismas características del defecto producido, tal como se han aclarado en el paso de *observación*. Aunque la reproducción intencional es un método efectivo para someter a prueba una hipótesis, hay ocasiones en las cuales no es permisible hacerlo debido a razones humanas, sociales o prácticas (tiempo o economía). En estos casos, debemos ser más cuidadosos en la realización de los pasos de *observación* y de *análisis*.

10.4 ACCIÓN

Realice acciones para eliminar las principales causas

Actividades

1) Debe hacerse una distinción estricta entre las acciones realizadas para solucionar fenómenos (remedio inmediato) y las acciones realizadas para eliminar los factores causales (prevención de recurrencia).
2) Cerciórese de que las acciones no producen otros problemas (efectos secundarios). Si lo hacen, adopte otras acciones, o diseñe medidas para los efectos secundarios.
3) Diseñe varias propuestas diferentes de acción, examine las ventajas y las desventajas de cada una y seleccione aquéllas que sean aceptadas por las personas involucradas.



Notas

1) Hay dos tipos de acción. Una es la acción para manejar fenómenos (resultados), mientras que la otra es para prevenir que el factor que causa el resultado ocurra de nuevo. Si producimos un producto defectuoso, reparamos el producto. Aun si tenemos éxito en la reparación, ésta no evitará que el defecto vuelva a presentarse. La manera ideal de solucionar un problema es evitar que se repita, adoptando medidas que eliminen la causa del problema. No deben confundirse los dos tipos de acción. Adopte siempre procedimientos que eliminen las causas.
2) Las acciones con frecuencia causan otros problemas. Son semejantes al uso de un tratamiento médico que cura una enfermedad, pero que tiene efectos secundarios y causa otra enfermedad al paciente. Para evitar los efectos secundarios, la acción tiene que evaluarse detenidamente y juzgarse desde un horizonte tan amplio como sea posible. También deben realizarse pruebas preparatorias (experimentos) sobre el método. Si se presentan efectos secundarios, considere la posibilidad de otra acción o de un remedio para los efectos secundarios.
3) Un aspecto importante en la selección de acciones es el hecho de poder asegurar la cooperación activa de todos los involucrados. Una acción que ataque un factor causal generará cambios en las prácticas de trabajo. Todos deben estar de acuerdo con la nueva práctica. Si existen varias acciones posibles, las ventajas y desventajas de cada una deben mirarse desde el punto de vista de las personas involucradas. Para la decisión final, si hay varias soluciones posibles que satisfagan igualmente bien las condiciones económicas y técnicas, es mejor seleccionar una democráticamente.

10.5 VERIFICACIÓN

Asegúrese de que el problema haya sido prevenido desde su raíz

Actividades

1) Compare los datos obtenidos sobre el problema (resultados indeseados en el tema), en el mismo formato (tablas, gráficas, esquemas) antes y después de realizadas las acciones.
2) Convierta el efecto en términos monetarios, y compare el resultado con el valor objetivo.
3) Haga una lista de cualquier otro efecto, bueno o malo.

Notas

1) En el paso de *verificación* nos preguntamos, "¿qué tan bien se ha prevenido la recurrencia?" Los datos que debemos usar para verificar la efectividad de las acciones son tomados antes y después de que se han [illegible] las acciones. En el paso de la *verificación* se hace una comparación de las situaciones antes y después de que se han realizado las acciones para determinar hasta qué punto se han reducido los efectos indeseables. El formato usado en esta comparación (tablas, gráficas, diagramas) debe ser el mismo antes y después de las acciones. Por ejemplo, si se usa un diagrama de Pareto para indicar la situación antes de la implementación de las acciones, entonces debe usarse un diagrama de Pareto para verificar la efectividad de esas acciones.
2) En la administración, es importante tratar de convertir los resultados de las acciones a términos monetarios. Pueden descubrirse datos de interés para la administración cuando se comparan las pérdidas antes y después de las acciones.
3) Cuando el resultado de las acciones no es tan satisfactorio como se esperaba, cerciórese de que las acciones planeadas se

han implementado precisamente de acuerdo con la decisión. Si los resultados no deseados continúan ocurriendo después de que se han realizado las acciones, ha fracasado la solución del problema, y es necesario regresar al paso de la *observación* y empezar de nuevo.

10.6 ESTANDARIZACIÓN

Elimine permanentemente la causa del problema

Actividades

1) Para el trabajo mejorado debe identificarse claramente: quién, cuándo, dónde, qué, por qué y cómo, y usarse como un estándar.
2) Las preparaciones y comunicaciones necesarias respecto a los estándares deben realizarse correctamente.
3) Deben implementarse la educación y el entrenamiento.
4) Debe diseñarse un sistema de responsabilidad para verificar si los estándares se están observando.

Notas

Las acciones correctivas deben estandarizarse para evitar que el problema recurra permanentemente. Hay dos razones principales para la estandarización. La primera es que sin estándares, las acciones tomadas para solucionar un problema gradualmente revertirán a las formas antiguas y eso llevará a una recurrencia del problema. La segunda es que sin estándares claros, es probable que el problema vuelva a presentarse cuando nuevas personas (nuevos empleados, transferencias, trabajadores de tiempo parcial) se involucren en el proceso. La estandarización no se logra solamente con documentos. Los estándares deben convertirse en parte de los pensamientos y de los hábitos de los trabajadores. Se requieren educación y entrenamiento para proporcionarles el conocimiento y las técnicas para implementar los estándares.



1) La estandarización es otra manera de expresar el quién, cuándo, dónde, qué, por qué y cómo de los procedimientos de trabajo*. Sólo mostrar el cómo puede llamarse un estándar en algunas ocasiones; un estándar puede considerarse satisfactorio si están presentes el quién, cuándo, dónde, qué y cómo (faltando el por qué), porque se puede comprender bien la forma de realizar un trabajo sin comprender el "por qué". Pero el "por qué" es indispensable para la persona que hace el trabajo. Hay muchos otros métodos, además del estándar, para realizar un trabajo y obtener resultados. Por tanto, es muy posible que el trabajador use un método no estándar si no conoce el por qué de la utilización del método estándar. Ésta es la razón por la cual el "por qué" debe incluirse en el estándar. Una vez que las personas han comprendido el "por qué", observarán cuidadosamente el estándar. La ruta de la calidad es una buena herramienta para comprender el "por qué". De esta manera, los estándares no pueden separarse de la ruta de la calidad que los produjo. Cuando se da entrenamiento y educación sobre los estándares, debe estudiarse también la ruta de la calidad relacionada con ellos.
2) La falta de una adecuada preparación y comunicación es una de las principales razones para la confusión cuando se introducen nuevos estándares. Llevar a la práctica nuevos estándares cambia las prácticas de trabajo y esto puede generar confusión, producida por errores triviales; y a veces se presentan problemas, particularmente en lugares de trabajo en donde se usa un sistema de división del trabajo, si en un lugar se están haciendo las cosas según el nuevo método y en otro lugar se están haciendo según el método antiguo.
3) Con frecuencia se necesitan la educación y el entrenamiento para lograr que los estándares se conviertan en hábitos. Si la empresa es negligente en este entrenamiento, independientemente de las bondades de los estándares en sí mismos, és-

* Llamados las 5W 1H por sus siglas en inglés que son *Who, When, Where, What, Why* y *How*.



tos no se utilizarán como se debería y es inevitable que vuelvan a presentarse problemas.

4) A veces se soluciona un problema, y el mismo vuelve a presentarse más tarde. La causa más importante de esto es que los estándares se observaron inicialmente pero eventualmente se relajaron. Debe organizarse un sistema de responsabilidades para verificar la observación constante de los estándares con el fin de evitar la recurrencia de los problemas.

10.7 CONCLUSIÓN

Revise el procedimiento seguido en la solución del problema y planee el trabajo futuro

Actividades

1) Haga una lista de los problemas que permanecen.
2) Planee qué hay que hacer para solucionar esos problemas.
3) Piense sobre lo que ha funcionado bien y lo que no ha funcionado en las actividades de mejoramiento.

Notas

1) Casi nunca se soluciona un problema perfectamente y la situación ideal casi nunca existe. No es bueno tratar de lograr la perfección ni continuar con las mismas actividades sobre el mismo tema por demasiado tiempo. Es importante delimitar las actividades cuando se alcanza la fecha límite original. Aun si no se ha logrado la meta, debe hacerse una lista de hasta dónde han progresado las actividades y de lo que no se ha logrado todavía.
2) Haga planes sobre lo que debe hacerse en el futuro con los problemas que subsisten. Los problemas importantes en esos planes deben ser tenidos en cuenta como temas en la siguiente ruta de la calidad.
3) Finalmente, debe dedicarse algo de reflexión a las actividades



realizadas para llevar a cabo la solución del problema. Esto ayudará a elevar la calidad de futuras acciones de mejoramiento. Siempre hay una diferencia entre la actividad que realmente se realizó y la que se comprendió intelectualmente que había de realizarse, y estas diferencias deben solucionarse una por una. Esta revisión debe hacerse aún si el problema se arregló con éxito, pero esta "práctica cerebral" debe realizarse con especial cuidado si ha pasado la fecha límite y el problema no se ha solucionado. Estos problemas pueden retomarse en la etapa siguiente de la ruta de la calidad.

Epílogo

HECHOS

Las causas generales de los problemas en las fábricas se deben a conocimientos erróneos y a operaciones incorrectas.

Para discernir lo erróneo y lo incorrecto debemos iniciar procesos para encontrar los hechos.

"Los hechos"

- Una expresión usada en exceso.
- Todo el mundo supone que sabe, pero en realidad nadie sabe.
  Podríamos recordar la historia de los hombres ciegos que tocaron un elefante, y cada uno dio una descripción diferente de lo que es un elefante.
  Uno toca la trompa y sólo habla de eso; otro toca la cola y describe al elefante como tal.
  Cada uno cree que su experiencia es la correcta.
  La gente con frecuencia narra historias de otros como si fueran de su propia experiencia.
- Las discusiones por sí solas no eliminan los problemas.
- Los hechos no siempre pueden describirse con palabras.
- Lo blanco puede resultar negro.
- Las discusiones no definirán si es blanco o es negro.



"Que hablen los hechos"

- Con actitud humilde, verifique cuidadosamente las cosas una por una.
  De todos modos, nos estamos entendiendo con algo difícil.
  Tiene un número infinito de características.
  Debemos entender que nuestro conocimiento y nuestra experiencia son limitadas y siempre imperfectas.
  Este reconocimiento hará que aparezcan los hechos.
- Llamamos experimentada a una persona que ha realizado una tarea por largo tiempo.
  Una persona experimentada tiene un gran conocimiento sobre esa tarea.
  Hay conocimiento correcto y hay conocimiento incorrecto.
  El problema es que la persona no sabe diferenciarlos.
  El verdadero experto es la persona experimentada que siempre está alimentando su conocimiento con hechos, que reflexiona sobre ese conocimiento y hace las correcciones.
  Infortunadamente, no todas las personas que tienen experiencia son necesariamente verdaderos expertos.
  Pueden convertirse en obstáculos portadores de superstición.
- Para encontrar el verdadero conocimiento debemos trabajar con diligencia.
  Es lo mismo que cuando subimos por un camino de montaña: debemos hacerlo paso a paso.
  Después de haber subido durante un tiempo, se dará cuenta repentinamente de que está muy por encima del punto de partida.

Apéndice

Tabla A.1 Tabla de la distribución normal

Ejemplo

El valor de ϵ para $k_{\epsilon} = 1.96$ se obtiene de la siguiente forma: parta de la primera columna de la derecha marcada 1.9*, pare en la columna marcada 6 en la primera hilera, y encontrará el valor 0.0250.

Tabla A.2 Coeficientes para la gráfica $\bar{x}-R$

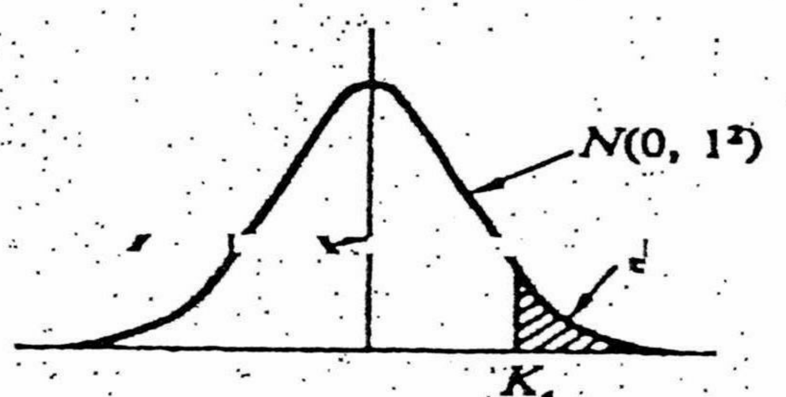
Tabla A.3 Puntos de porcentaje para la distribución t

1) Se usa esta tabla para obtener el valor t cuando se dan el grado de libertad (ϕ) y la probabilidad bilateral (α). Por ejemplo, cuando se dan $\phi=10$ y $\alpha = 0.05$, entonces $t = 2.228$.
2) En nuestra notación, el valor en la tabla se designa por $t(\phi, \alpha)$. De modo que $t(10, 0.05) = 2.228$.
3) Cuando $\phi = \infty$, los valores siguen los percentiles de la distribución normal estándar, es decir, $t(\infty, \alpha) = u(\alpha)$.



k_ϵ	*=0	1	2	3	4	5	6	7	8	9
0.0*	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1*	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2*	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3*	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4*	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5*	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6*	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7*	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8*	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9*	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0*	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1*	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2*	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3*	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4*	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5*	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6*	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7*	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8*	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9*	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0*	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1*	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2*	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3*	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4*	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5*	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6*	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7*	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8*	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9*	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0*	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Tabla A.1 Tabla de la distribución normal





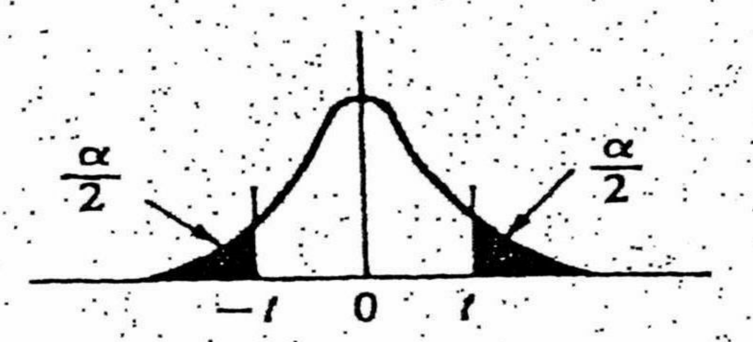
Tamaño de la muestra n	Gráfica $\bar{x}$	Gráfica R				
	A_2	d_2	$1\backslash d_2$	d_3	D_3	D_4
2	1.880	1.128	0.8862	0.853	—	3.267
3	1.023	1.693	0.5908	0.888	—	2.575
4	0.729	2.059	0.4857	0.880	—	2.282
5	0.577	2.326	0.4299	0.864	—	2.115
6	0.483	2.534	0.3946	0.848	—	2.004
7	0.419	2.704	0.3698	0.833	0.076	1.924
8	0.373	2.847	0.3512	0.820	0.136	1.864
9	0.337	2.970	0.3367	0.808	0.184	1.816
10	0.308	3.078	0.3249	0.797	0.223	1.777

Nota: El símbolo "—" en la columna D_3 significa que no se ha considerado el límite de control inferior.

Tabla A.2 Coeficientes para la gráfica $\bar{x} - R$

ϕ \ α	0.10	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2,845
25	1.708	2.060	2.485	2.787
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.326	2.576

Tabla A.3 Puntos de porcentaje para la distribución t



Respuestas a los ejercicios

2.1

Lo que dicen los trabajadores o los supervisores puede ser la verdad, pero deben confrontarse los hechos con la ayuda de los datos, los cuales deben recogerse de tal manera que muestre qué tipo de desperfecto prevalece y cuál es la causa del desperfecto. Teniendo esto en cuenta, el problema debe afrontarse utilizando los siguientes pasos:

1) Estratifique los tipos de desperfectos y averigüe cuál de los tipos es el mayor problema.
2) Haga una lista de las posibles causas de esos desperfectos, por ejemplo, materiales, partes, las mismas máquinas, herramientas, operarios, el método de medición, etc.
3) Recoja la información de tal manera que sea posible aislar el efecto de cada causa. Debido a que en este ejemplo hay dos operarios y dos máquinas, al menos estas dos posibles causas deben examinarse por estratificación.
4) Cuando las causas no pueden estratificarse con claridad, se han de llevar registros. Esto debe hacerse, por ejemplo, cuando cambian los materiales o los métodos de operación.
5) La información debe analizarse con la ayuda de diagramas de Pareto, hojas de verificación, diagramas de dispersión, etc.
6) Una vez se haya aclarado cuál es la causa de los productos



defectuosos, deben tomarse acciones correctivas y sus resultados deben observarse cuidadosamente.

3.1

1) El diagrama general de Pareto que se muestra en la figura indica que el defecto más numeroso es la *tensión*. Esto también es verdad en la mayoría de los diagramas por ítem, de modo que el defecto que se debe atacar en primer lugar es la tensión.

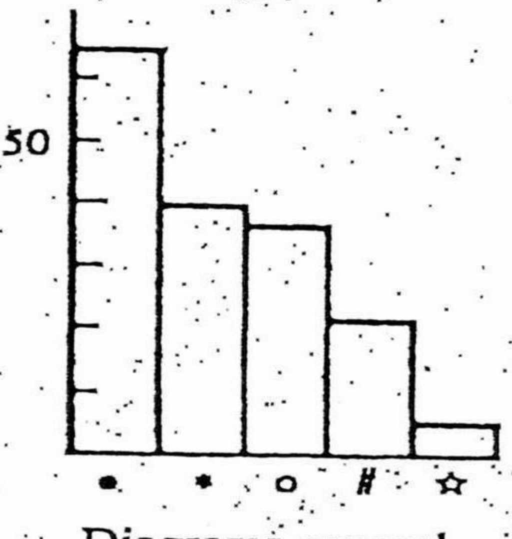
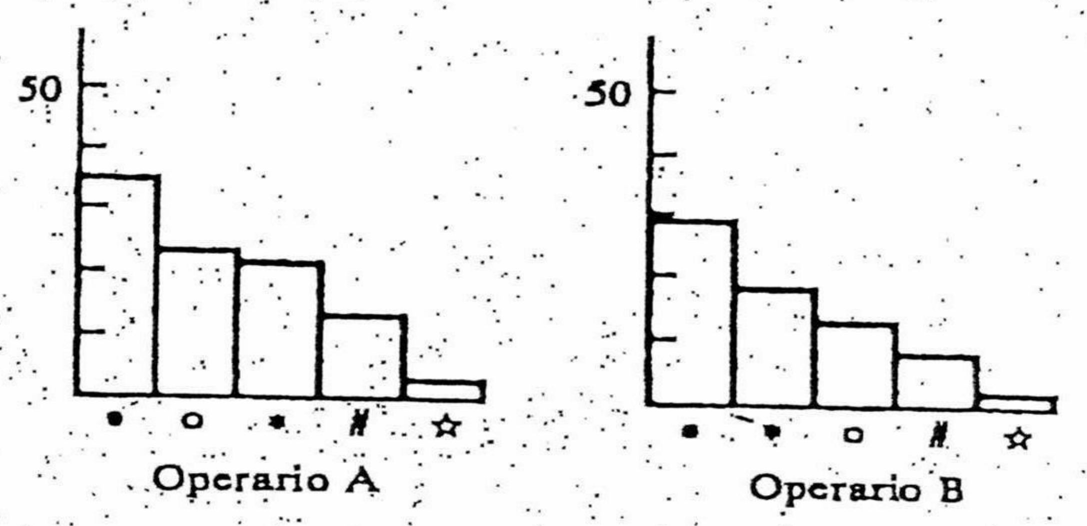
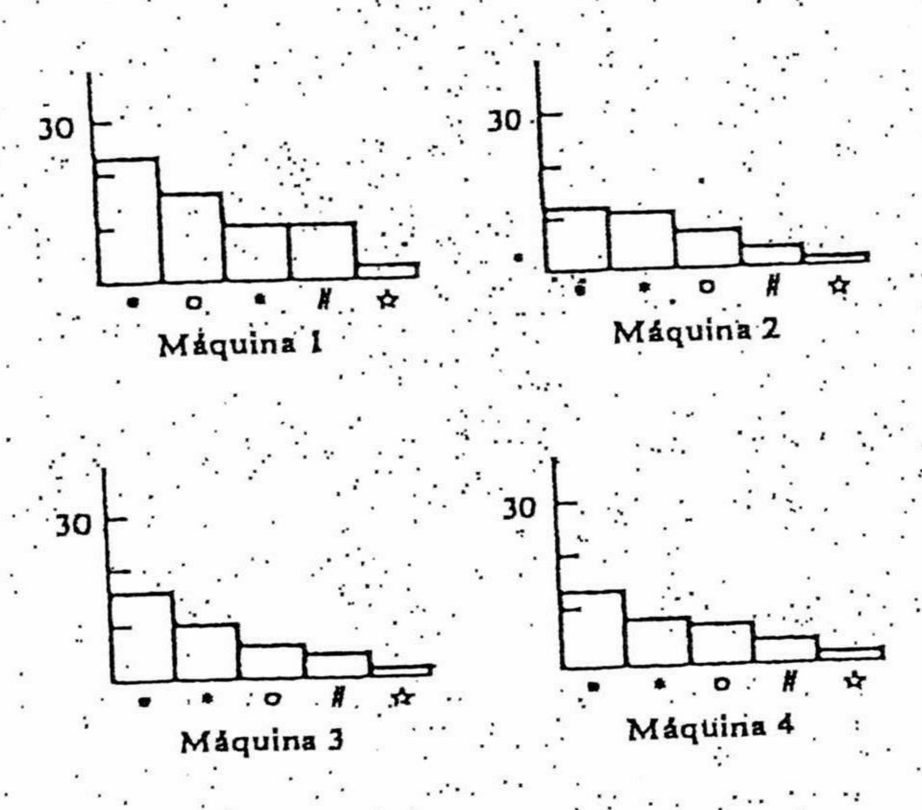


Diagrama general

2) Cuando comparamos los dos operarios *A* y *B*, es evidente que *A* produce un mayor número de productos defectuosos que *B*. Debe existir alguna razón para esto, de manera que debemos tratar de analizar esta situación más a fondo.



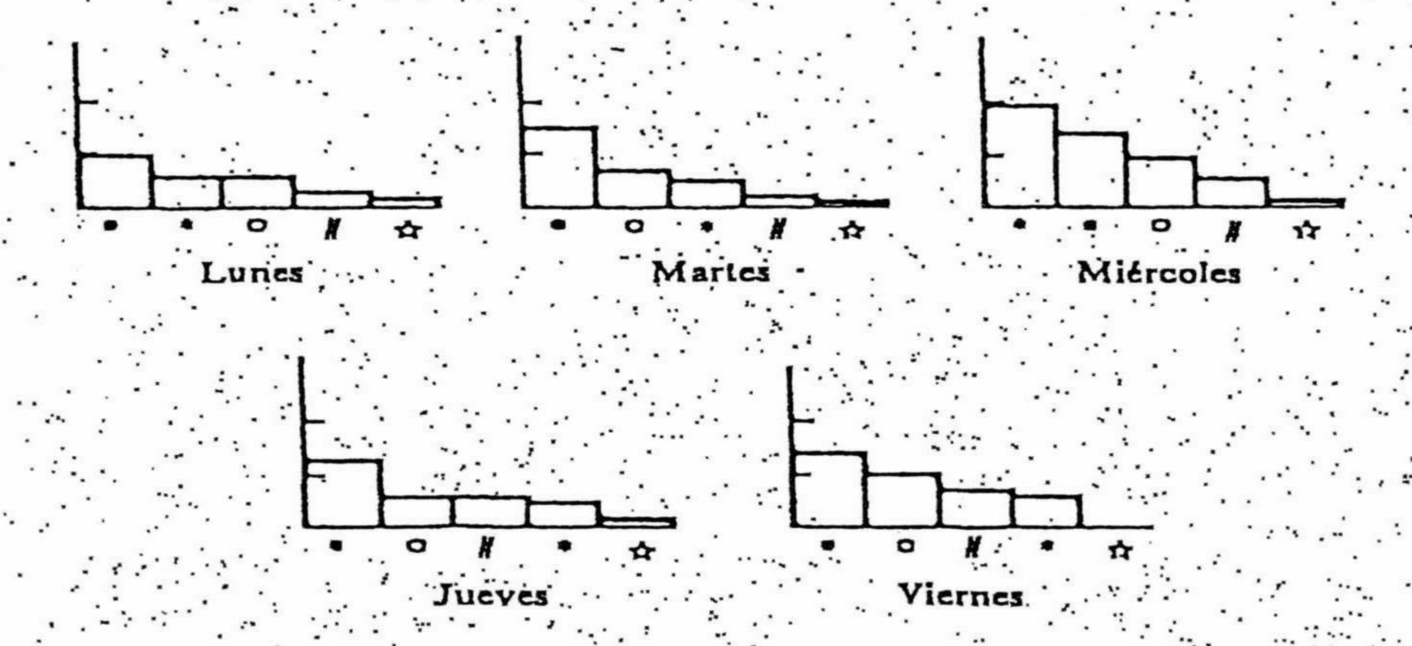
3) Cuando comparamos las cuatro máquinas, 1, 2, 3 y 4, vemos que en general la máquina 1 produce más defectos que cualquiera otra. Pero esta máquina es manejada solamente por el operario *A*, y no por el operario *B*. De modo que parece que el mayor número de ítems defectuosos producidos por la máquina 1 tiene alguna influencia sobre el hecho de que la fracción general de ítems defectuosos producida por el operario *A* sea mayor que la producida por el operario *B*. El hecho de que el operario *A* también trabaje con la máquina 2 da más fuerza a esta observación, pero no existe una disparidad obvia entre las clasificaciones de defectos producidos por esta máquina y las máquinas 3 y 4. Parece que la máquina 1 es la causa del problema, y por tanto debemos investigar esto más a fondo.



4) Si comparamos el número de defectos producidos en diferentes días de la semana, vemos que ocurren más defectos el miércoles que cualquier otro día. Si luego miramos la clasificación de defectos para ese día, es evidente que se concentra en los *rayones*, mientras que los otros tipos de defectos ocurren más o menos en los mismos números. Al verificar la



tabla con los datos brutos (tabla 3.3), vemos que los dos operarios y las cuatro máquinas producen más rayones los miércoles que cualquier otro día. De esto debe existir alguna razón y debemos investigarla más.



Resumen

Los cursos de acción sugeridos por este análisis pueden resumirse de la siguiente manera:

a) Si se revisa la máquina 1, debe ser posible disminuir ligeramente el número de defectos en todas las categorías.
b) Si se identifica la causa del número anormal de rayones que ocurre los miércoles, debe ser posible reducir el número de este tipo de defectos.
c) Si se toma como meta de la investigación la tensión, que es el más prevalente de los defectos, debe ser posible disminuir el número general de ítems defectuosos.

Afrontando el problema de esta manera sistemática, será posible lograr una disminución drástica del número general de ítems defectuosos.

4.1

1) Un diagrama de causa-efecto de errores de mecanografía. El diagrama a continuación es una solución a este problema. Recuerde, sin embargo, que ésta no es de ninguna manera la única solución posible. Los diagramas de causa-efecto son de tal naturaleza que para cualquier problema hay una gran variedad de soluciones alternas, y no puede decirse que una de ellas sea la única correcta.

2) y 3) deben considerarse individualmente.

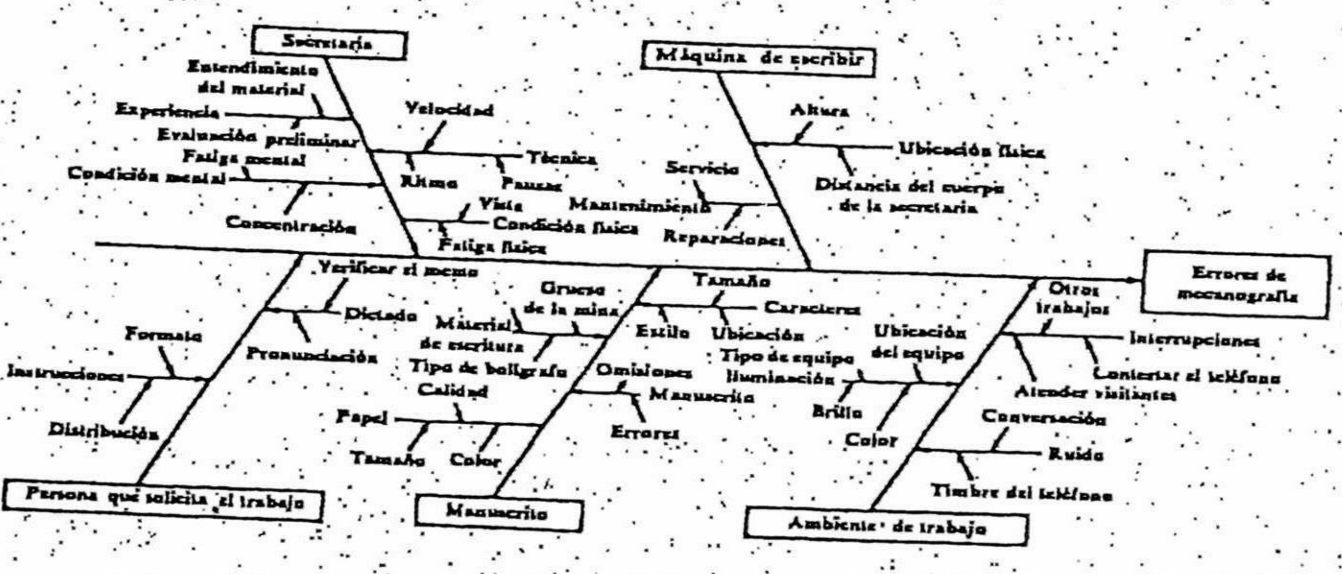


Diagrama de causa-efecto de los errores de mecanografía

5.1

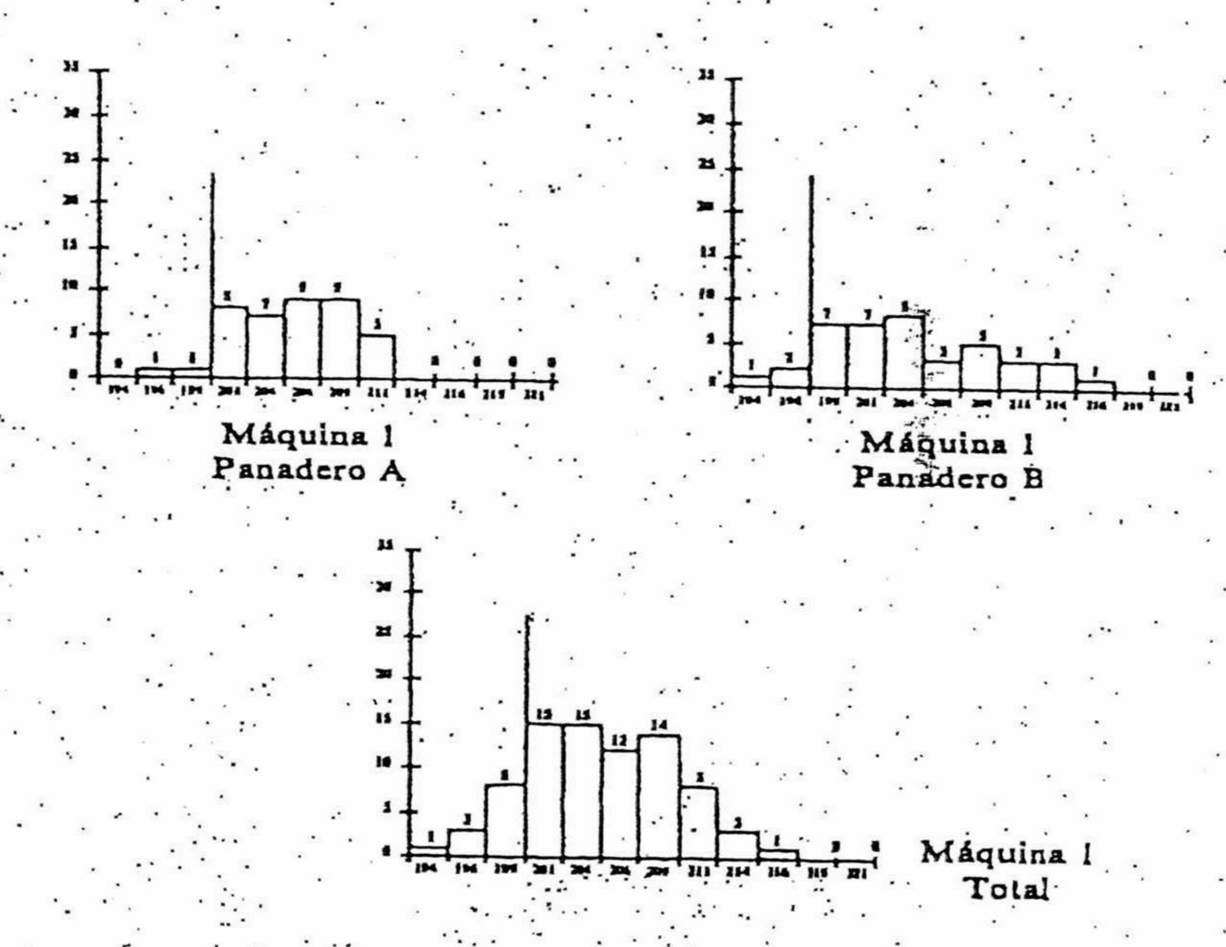
1) Para las diferentes clases de histogramas y sus estadígrafos, vea la figura y la tabla.
2) Comentarios:
   a) Todos los panes que están bajos de peso son producidos por la máquina 1, y a partir de los histogramas para esta máquina, vemos que ellos son de doble pico. Así, la desviación estándar, *s*, es grande y es probable que la no conformidad la esté causando el pico de la izquierda. Esta situación debe analizarse cuidadosamente, y si los dos picos pueden clasificarse separadamente, debe ser posible reducir la variación en los datos y disminuir la tasa de no conformidad.

b) En el histograma del panadero *A* usando la máquina 2, hay un valor que se aparta del resto de los datos. Es muy posible que este valor sea espurio, y deba verificarse.

c) Las medias de todos estos histogramas son menores que el valor central de los límites de especificación, 212.5. Si se aclara la causa para esto y se encuentra la manera de acercar las medias a este valor central, la tasa de no conformidad debe disminuir. Esto no necesariamente mejorará la variación, que es todavía demasiado grande, pero si las acciones sugeridas en a) y b) tienen éxito, la variación debe reducirse a un valor suficientemente bajo.

Si se adoptan todas las tres medidas anteriores, el proceso se ajustará a la especificación.

Histogramas



Máquina 1
Panadero A

Máquina 1
Panadero B

Máquina 1
Total

Nota: Todo lo que figure al lado izquierdo de la raya está fuera de la especificación

Máquina 2
Panadero A

35
30
25
20
15
10
5
0

0 0 0 0 4 4 10 10 6 5 0 1

194 196 199 201 204 206 209 211 214 216 219 221

Máquina 2
Panadero B

35
30
25
20
15
10
5
0

0 0 0 0 1 7 11 10 10 1 0 0

194 196 199 201 204 206 209 211 214 216 219 221

Máquina 2
Total

35
30
25
20
15
10
5
0

0 0 0 0 5 11 21 20 16 6 0 1

194 196 199 201 204 206 209 211 214 216 219 221



Panadero A
Total

194	196	199	201	204	206	209	211	214	216	219	221
0	1	1	8	11	13	19	15	6	5	0	1

Panadero B
Total

194	196	199	201	204	206	209	211	214	216	219	221
1	2	7	7	9	10	16	13	13	2	0	0

Total

194	196	199	201	204	206	209	211	214	216	219	221
1	3	8	15	20	23	35	28	19	7	0	1



	Panadero *A*	Panadero *B*	Total
Máquina 1	$n = 40$ $\bar{x} = 205.37$ $s = 3.75$	$n = 40$ $\bar{x} = 204.40$ $s = 5.38$	$n = 80$ $\bar{x} = 204.88$ $s = 4.63$
Máquina 2	$n = 40$ $\bar{x} = 210.66$ $s = 4.15$	$n = 40$ $\bar{x} = 210.34$ $s = 2.89$	$n = 80$ $\bar{x} = 210.50$ $s = 3.56$
Total	$n = 80$ $\bar{x} = 208.01$ $s = 4.75$	$n = 80$ $\bar{x} = 207.37$ $s = 5.23$	$n = 160$ $\bar{x} = 207.69$ $s = 4.99$

Valores de n, $\bar{x}$ y s para cada histograma

6.1

1)

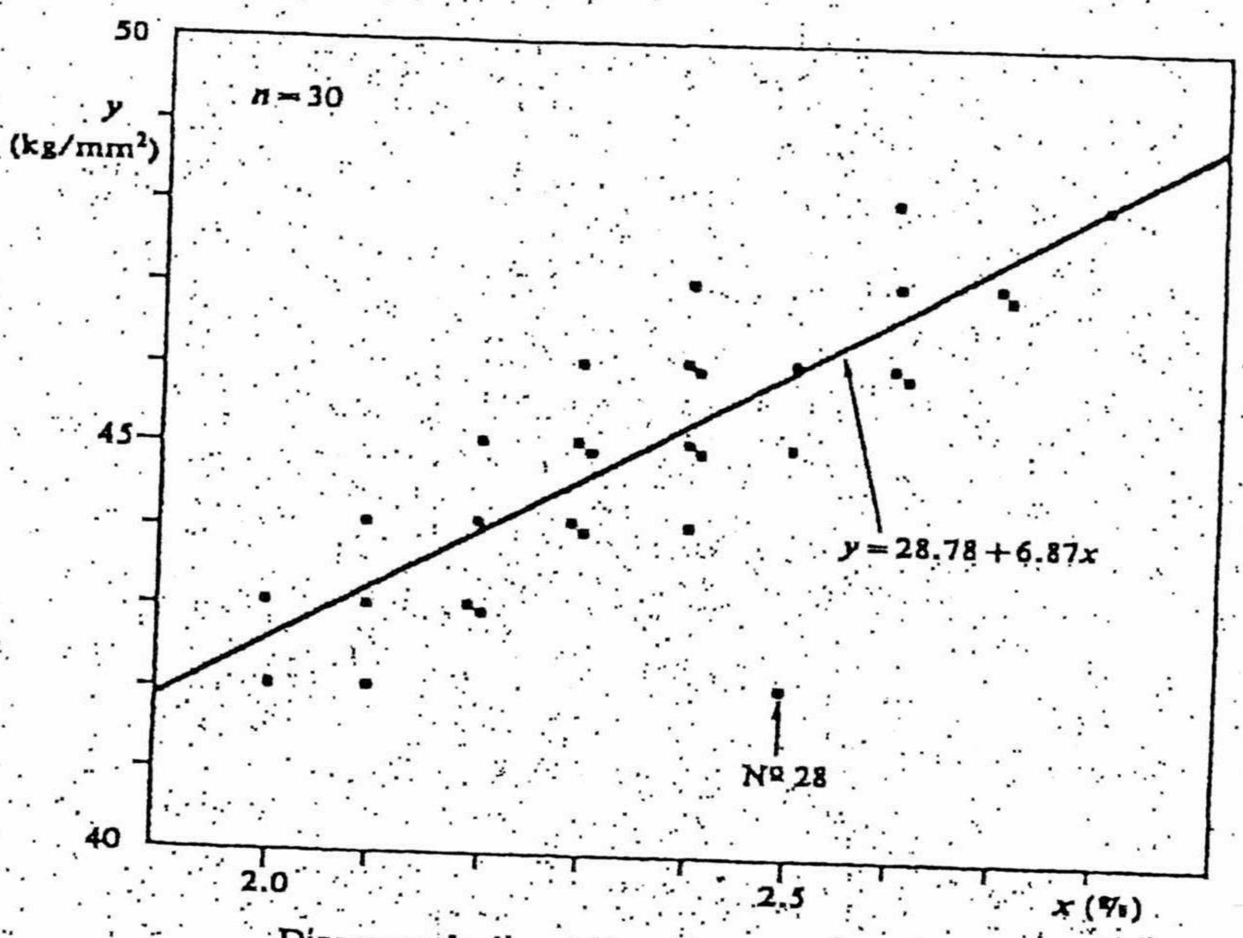


Diagrama de dispersión y línea de regresión



a) Existe una correlación positiva entre x y y.
b) Hay un punto (No. 28, $x = 2.5\%$, $y = 42$ kg/mm^2) que es distinto de los otros puntos. Debe investigarse por qué estos datos incluyen un valor extremo.

2) $r = 0.789$.
$r = 0.877$, cuando se elimina el dato No. 28.

3) $y = 29.58 + 6.48\,x$.
$y = 28.78 + 6.87\,x$, cuando se elimina el dato No. 28.

7.1

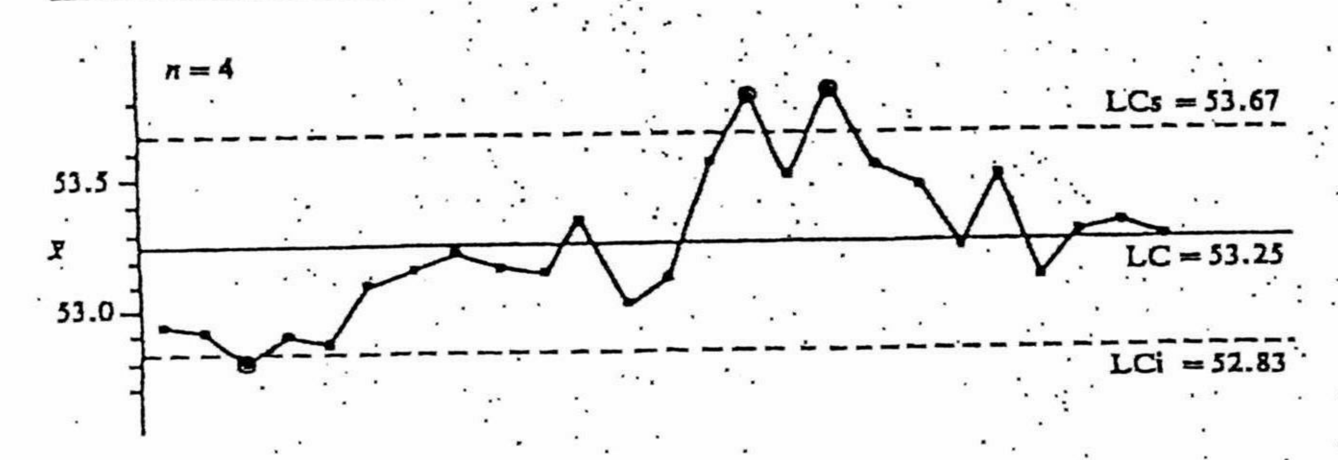
1) $\bar{x} - R$
2) pn
3) c
4) $x - Rs$
5) p
6) $\bar{x} - R$
7) u

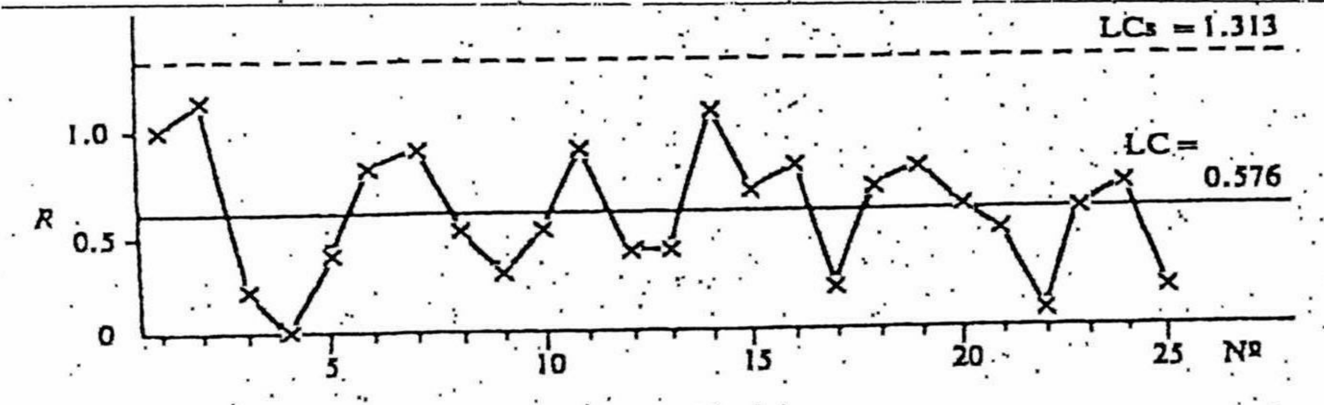
7.2

$\bar{x} = 53.25$
$\bar{R} = 0.576$
Líneas de control

Gráfica $\bar{x}$ LCs $= \bar{\bar{x}} + A2\ \bar{R} = 53.25 + 0.729 \times 0.576 = 53.67$
LC $= \bar{\bar{x}} = 53.25$
LCi $= 52.83$

Gráfica R LCs $= D_4\ \bar{R} = 2.28 \times 0.576 = 1.313$
LC $= \bar{R} = 0.576$
LCi $=$ __________ (no se tiene en cuenta)





La gráfica *R* muestra un estado de control. Sin embargo, la gráfica $\bar{x}$ no muestra un estado de control:

1) Hay tres puntos por fuera de los límites de control, los números 3, 15 y 17.
2) Hay una racha de una longitud de 10 (Nos. 1-10).

Por tanto, este proceso no está en un estado estable. La media del proceso está en un nivel más bajo antes del 13 de noviembre, y de nuevo sube y baja. Debemos encontrar la causa del cambio en la media del proceso.

7.3

1) Correcto.
2) Correcto.
3) Debemos también verificar si los puntos forman algún tipo de patrón particular o no.



4) Correcto.
5) Cuando un punto cae por fuera del LCi, debemos considerar que el proceso está fuera de control. Podemos conseguir información útil si investigamos la razón por la cual la característica de control está por debajo del LCs.
6) Un proceso en el estado de control puede producir productos defectuosos.
7) Cuando una gráfica de control muestra un estado no controlado, en todo caso debemos tratar de encontrar la causa.
8) Correcto.
9) Es posible que un subgrupo se componga de datos heterogéneos. Debemos verificar el contenido de la variación dentro del subgrupo, y ensayar otras manera de formar los subgrupos.

8.1

Suma y diferencia de *A* y *B*

No.	*A* + *B*	*A* − *B*
1	12.35	1.55
2	11.20	2.30
3	11.90	2.60
4	11.05	1.95
5	12.90	3.00
6	11.60	3.80
7	11.10	2.60
8	11.45	3.55
9	11.85	2.25
10	12.80	3.00

Estadísticas de *A*, *B*, *A* + *B*, y *A* − *B*

	A	*B*	*A* + *B*	*A* − *B*
Total	72.40	45.80	118.20	26.6
Suma de cuadrados	2.279	1.951	4.096	4.364
Varianza	0.253	0.217	0.455	0.485

8.2

Sea el actual valor c, y los pesos del contenido y el contenedor sean x, y respectivamente, entonces,

$$x + y = c + e,$$

donde e es el error de la máquina al pesar.

El peso del contenido está dado por

$$x = c + e - y$$

c es una constante, e y y son independientes, de modo que la varianza de x se convierte en la suma de la varianza de y y e, es decir

$$\sigma_x^2 = \sigma_1^2 + \sigma_2^2$$

9.1

1) Una población es algo sobre lo cual actuamos, y una muestra es una pequeña parte de la población. Basados en la muestra, podemos estimar las características de toda la población.
2) Un parámetro de la población es una constante que caracteriza a la población y una estadística es una función de una muestra de observaciones obtenidas de la población. Basados en el valor de la estadística, podemos hacer una inferencia sobre el valor correspondiente del parámetro de la población.
3) Cuando una hipótesis se somete a prueba, pueden cometerse dos tipos de error. Un error del primer tipo (error tipo I) es el de rechazar la verdadera hipótesis, y un error del segundo tipo (error tipo II) es el de aceptar una hipótesis falsa.
4) Un nivel de significación es la probabilidad de rechazar incorrectamente una hipótesis, y un nivel de confianza es la probabilidad de que los límites de confianza incluyan el verdadero valor del parámetro estimado.



9.2

No.	$\overline{x}$	R	V	s	u	t
11	50.4	5	3.80	1.95	0.45	0.46
12	51.8	4	2.20	1.48	2.01	2.72
13	50.0	4	2.50	1.58	0.00	0.00
14	50.0	4	2.50	1.58	0.00	0.00
15	50.4	5	3.80	1.95	0.45	0.46
16	50.8	8	8.20	2.86	0.89	0.62
17	50.6	5	5.30	2.30	0.67	0.58
18	50.6	4	2.30	1.52	0.67	0.88
19	49.6	6	4.80	2.19	−0.45	−0.41
20	50.6	5	3.80	1.95	0.67	0.69

9.3

1) 1. Hipótesis, nivel de significación

H_0: $\mu_1 - \mu_2$

H_1 : $\mu_1 > \mu_2$ ($\alpha = 0.05$)

2. Estadísticas

$\sum x_{1i} = 77.5$

$\sum x_{1i}^2 = 1222.33$

$\sum x_{2i} = 68.1$

$\sum x_{2i}^2 = 942.23$

$$\overline{x}_1 = \frac{77.5}{5} = 15.50$$

$$\overline{x}_2 = \frac{68.1}{5} = 13.62$$

$$S_1 = 1222.33 - \frac{77.5^2}{5} = 21.80$$

$$S_2 = 942.23 - \frac{68.1^2}{5} = 14.71$$

$$s=\sqrt{\frac{21.08+14.71}{5+5-2}}=\sqrt{\frac{35.79}{8}}=\sqrt{4.474}=2.115$$

$$t_0=\frac{15.50-13.62}{2.115\cdot\sqrt{\frac{1}{5}+\frac{1}{5}}}=\frac{1.88}{1.3376}=1.405$$

3. Prueba
   $t\,(8,\,0.10) = 1.860$
   $t_0 = 1.405 < 1.860 = t\,(8,\,0.10)$
   $\longrightarrow$ acepte H_0.

4. Conclusión
   No podemos concluir, al nivel de significación del 5%, que el tratamiento aumente la resistencia de abrasión.

2)

$$\widehat{\mu_1-\mu_2}=\bar{x}_1-\bar{x}_2=1.88$$

$$\mu_1-\mu_2=\bar{x}_1-\bar{x}_2\pm t(n_1+n_2-2,0.05)\cdot s\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$=1.88\pm 2.306\cdot 2.115\cdot\sqrt{\frac{1}{5}+\frac{1}{5}}$$

$$=1.88\pm 3.08$$

$$=-1.20\sim 4.96$$

3) La comparación debe hacerse entre especímenes tan semejantes como sea posible. Para hacerlo, es mejor cortar las piezas de prueba por la mitad, tratar una mitad y la otra no. El análisis se realiza para determinar si hay o no una diferencia en el interior de los especímenes pareados.

